

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2003, 11:57:49 ; Search time 10.9842 Seconds  
(without alignments)  
4722.097 Million cell updates/sec

Title: US-09-970-318-2  
Perfect score: 1092  
Sequence: 1 FLAQRVCIPNDAGTADNRV.....YTTETAVALRHRYYAVDEIT 201

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL.23.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1092	100.0	1546	4	O75445
2	1086	99.5	1546	4	O75445
3	759	69.5	1461	11	Q9JLFP3
4	728	66.7	1512	11	Q8K3K1
5	294	26.9	214	11	Q9D128
6	186	17.0	569	13	O57339
7	185	16.9	603	13	O42140
8	181	16.6	604	11	Q92429
9	175	16.0	670	5	O95YK2
10	173.5	15.9	3695	4	Q8TDF8
11	163.5	15.0	600	5	O9NL27
12	163.5	15.0	650	5	O9NL28
13	161	14.7	3704	5	P91904
14	160.5	14.7	602	13	O42203
15	160	14.7	529	4	Q8N2D6
16	159.5	14.6	555	5	Q9NFW6

17	159.5	14.6	1785	13	O8JHV7
18	149	13.6	1593	13	O8JHV8
19	142.5	13.0	1792	13	O57484
20	141.5	13.0	1486	4	O14637
21	140.5	12.9	667	5	O9VY25
22	139	12.7	628	11	O9J133
23	138	12.6	610	5	O96659
24	138	12.6	3712	5	O9VRW0
25	136.5	12.5	628	4	O9HB63
26	136.5	12.5	1623	5	O9U3U7
27	136.5	12.5	1808	5	O44565
28	135.5	12.4	628	4	O982P1
29	135	12.4	1827	13	O8JHV6
30	133.5	12.2	3102	5	O45614
31	131	12.0	168	13	O91435
32	131	12.0	761	4	O9UHI2
33	131	12.0	1631	4	O9Y6U6
34	130.5	12.0	1799	11	O8R0Y0
35	117.5	10.8	695	11	O8C9J2
36	116.5	10.7	530	11	O8R4F2
37	116.5	10.7	530	11	O8VIP8
38	116.5	10.7	564	11	O8VIP6
39	116.5	10.7	589	11	O8R4F1
40	116.5	10.7	589	11	O8VIP7
41	115.5	10.6	530	4	O96CW9
42	115.5	10.6	549	4	O96JH0
43	103.5	9.5	364	4	O9Y212
44	103.5	9.5	438	4	O8N633
45	102	9.3	580	11	O9QY49

## ALIGNMENTS

### RESULT 1

O75445	O75445	PRELIMINARY;	PRT; 1546 AA.
ID	O75445	PRELIMINARY;	PRT; 1546 AA.
AC	O75445		
DT	01-NOV-1998	(Tremblrel. 08, Created)	
DT	01-JUN-2002	(Tremblrel. 21, Last sequence update)	
DT	01-MAR-2003	(Tremblrel. 23, Last annotation update)	
DE	Usher syndrome type IIA protein.		
GN	USH2A.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
EX	MEDLINE=98288356; PubMed=9624053;		
RA	Eudy J.D., Weston M.D., Yao S.F., Hoover D.M., Rehm H.L., Ahmad I.,		
RA	Ma-Edmonds M., Yan D., Cheng J.J., Beisel K.W., Ayuso C., Cremers C.,		
RA	Davenport S., Moller C., Talmadge C.B., Tamayo M., Swaroop A.,		
RA	Morton C.C., Kimberling W.J., Sumegi J.;		
RT	"Mutation of a gene encoding a protein with extracellular matrix		
RT	motifs in Usher syndrome type IIA.";		
RL	Science 280:1753-1757(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Eudy J.D., Yao S.F., Cheng J.J., Weston M.D., Sumegi J.;		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF055580; AAC23748.2;		
DR	HSSP; P02468; IKLO.		
DR	Genew; HGNC:12601; USH2A.		
DR	InterPro; IPR006209; EGF like.		
DR	InterPro; IPR003961; FN III.		
DR	InterPro; IPR006558; LamG like.		
DR	InterPro; IPR002049; Laminin_EGF.		
DR	InterPro; IPR001886; LamNT.		
DR	Pfam; PF00041; fn3; 4.		
DR	Pfam; PF00053; laminin_EGF; 9.		
DR	Pfam; PF00055; laminin_Nterm; 1.		
DR	PRINTS; PR00011; EGF_LAMININ.		

O8JHV7 brachydanio  
O8JHV8 brachydanio  
O57484 gallus gall  
O14637 homo sapien  
O9VY25 drosophila  
O9J133 mus musculus  
O96659 hirudo medi  
O9VRW0 drosophila  
O9HB63 homo sapien  
O9U3U7 anopheles g  
O44565 caenorhabdi  
O9BZP1 homo sapien  
O8JHV6 brachydanio  
O45614 caenorhabdi  
O91435 gallus gall  
O9UHI2 homo sapien  
O9Y6U6 homo sapien  
O8R0Y0 mus musculus  
O8C9J2 mus musculus  
O8R4F2 mus musculus  
O8VIP8 mus musculus  
O8VIP6 mus musculus  
O8R4F1 mus musculus  
O8VIP7 mus musculus  
O96CW9 homo sapien  
O96JH0 homo sapien  
O9Y212 homo sapien  
O8N633 homo sapien  
O9QY49 mus musculus

```

DR SMART; SM00180; EGF Lam; 10.
DR SMART; SM00060; FN3; 4.
DR SMART; SM00560; LamGL; 1.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF 1; 7.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
KW Laminin EGF-like domain.
SQ SEQUENCE 1546 AA; 170985 MW; 49CD0A95A614959F CRC64;

Query Match 100.0%; Score 1092; DB 4; Length 1546;
Best Local Similarity 100.0%; Pred. No. 1.8e-97;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLAQRVCIPNDAGDTADNRVSRNLNPEAHPLSPVNDNDVGTSMVSNVFTNITQLNQGVTTIS 60
DB 313 PLAQRVCIPNDAGDTADNRVSRNLNPEAHPLSPVNDNDVGTSMVSNVFTNITQLNQGVTTIS 372
QY 61 VDLENGQYQVFIITQFPSPQPTTEIRIQKENSLEDWEDWQYFARNCGAFGMKNGDLEK 120
DB 373 VDLENGQYQVFIITQFPSPQPTTEIRIQKENSLEDWEDWQYFARNCGAFGMKNGDLEK 432
QY 121 PDSVNCQLQSNFTPYSGNVTFSILTPGPNRYRFGYNNFYNTPSLQESVKATQIRFHFHQ 180
DB 433 PDSVNCQLQSNFTPYSGNVTFSILTPGPNRYRFGYNNFYNTPSLQESVKATQIRFHFHQ 492
QY 181 YTTTAVNLRHRYAVDEIT 201
DB 493 YTTTAVNLRHRYAVDEIT 513

RESULT 2
Q9NS27 Q9NS27 PRELIMINARY; PRT; 1546 AA.
AC Q9NS27;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Usher syndrome type IIA protein.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20206315; PubMed=10729113;
RA Weston M.D., Eudy J.D., Fugita S., Yao S.-F., Usami S., Cremers C.,
RA Greenburg J., Ramesar R., Martini A., Moller C., Smith R.J.,
RA Sumegi J., Kimberling W.J.;
RT "Genomic structure and identification of novel mutations in usherin,
RT the gene responsible for Usher syndrome type IIA.";
RL Am. J. Hum. Genet. 66:1199-1210(2000).
DR EMBL; AF091889; AAF75819.1; JOINED.
DR EMBL; AF091873; AAF75819.1; JOINED.
DR EMBL; AF091874; AAF75819.1; JOINED.
DR EMBL; AF091875; AAF75819.1; JOINED.
DR EMBL; AF091876; AAF75819.1; JOINED.
DR EMBL; AF091877; AAF75819.1; JOINED.
DR EMBL; AF091878; AAF75819.1; JOINED.
DR EMBL; AF091879; AAF75819.1; JOINED.
DR EMBL; AF091880; AAF75819.1; JOINED.
DR EMBL; AF091881; AAF75819.1; JOINED.
DR EMBL; AF091882; AAF75819.1; JOINED.
DR EMBL; AF091883; AAF75819.1; JOINED.
DR EMBL; AF091884; AAF75819.1; JOINED.
DR EMBL; AF091885; AAF75819.1; JOINED.
DR EMBL; AF091886; AAF75819.1; JOINED.
DR EMBL; AF091887; AAF75819.1; JOINED.
DR EMBL; AF091888; AAF75819.1; JOINED.
DR HSSP; P02468; IXLQ.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR006558; LamG like.
DR InterPro; IPR002049; Laminin_EGF.

```

```

DR InterPro; IPR001886; LamNT.
DR Pfam; PF00041; fn3; 4.
DR Pfam; PF00053; laminin_EGF; 10.
DR Pfam; PF00055; laminin_Nterm; 1.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00180; EGF_Lam; 10.
DR SMART; SM00060; FN3; 4.
DR SMART; SM00560; LamGL; 1.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF 1; 7.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
KW Laminin EGF-like domain.
SQ SEQUENCE 1546 AA; 171001 MW; 64BD435996C1E17B CRC64;

Query Match 99.5%; Score 1086; DB 4; Length 1546;
Best Local Similarity 99.5%; Pred. No. 6.8e-97;
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PLAQRVCIPNDAGDTADNRVSRNLNPEAHPLSPVNDNDVGTSMVSNVFTNITQLNQGVTTIS 60
DB 313 PLAQRVCIPNDAGDTADNRVSRNLNPEAHPLSPVNDNDVGTSMVSNVFTNITQLNQGVTTIS 372
QY 61 VDLENGQYQVFIITQFPSPQPTTEIRIQKENSLEDWEDWQYFARNCGAFGMKNGDLEK 120
DB 373 VDLENGQYQVFIITQFPSPQPTTEIRIQKENSLEDWEDWQYFARNCGAFGMKNGDLEK 432
QY 121 PDSVNCQLQSNFTPYSGNVTFSILTPGPNRYRFGYNNFYNTPSLQESVKATQIRFHFHQ 180
DB 433 PDSVNCQLQSNFTPYSGNVTFSILTPGPNRYRFGYNNFYNTPSLQESVKATQIRFHFHQ 492
QY 181 YTTTAVNLRHRYAVDEIT 201
DB 493 YTTTAVNLRHRYAVDEIT 513

RESULT 3
Q9JLP3 Q9JLP3 PRELIMINARY; PRT; 1461 AA.
AC Q9JLP3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative extracellular matrix protein MUSH2A.
GN MUSH2A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Eudy J.D., Talmadge C.B., Weston M.D., Yao S.-F., Cosgrove D.,
RA Ahmad I., Kimberling W., Sumegi J.;
RT "Isolation and Characterization of the Murine Homolog of the Usher
RT Syndrome Type 2A Gene.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF151717; AAF70550.1; -.
DR HSSP; P02468; 1TLE.
DR MGD; MGI:1341292; Ush2a.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001886; LamNT.
DR Pfam; PF00041; fn3; 4.
DR Pfam; PF00053; laminin_EGF; 10.
DR Pfam; PF00055; laminin_Nterm; 1.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00180; EGF_Lam; 8.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF 1; 7.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
KW Laminin EGF-like domain; Matrix protein.
SQ SEQUENCE 1461 AA; 160989 MW; 114436D877C38A15 CRC64;

```

Query Match 69.5%; Score 759; DB 11; Length 1461;  
 Best Local Similarity 68.5%; Pred. No. 5.4e-65;  
 Matches 137; Conservative 25; Mismatches 38; Indels 0; Gaps 0;

QY 1 PLAQRVCIPNDAGTADNRVSRNLNPEAHPLSFVNDNDVGTWSVSNVFTNTQLNQGVITIS 60  
 DB 310 PSVQVCIPNGAGDTPEHRMSRLNPEAHPLSFVNDNDVGTWSVSNVFTNTQLNQGVITIS 369  
 QY 61 VDLENGQYQVFIIOFFSPQTEIRIQKENSLEDWEDWQYFARNCGAFGKMGNGDLEK 120  
 DB 370 IDLENGQYQVFIIOFFSPQTEIRIQKENSLEDWEDWQYFARNCGAFGKMGNGDLEK 429  
 QY 121 PDSVNCQLQSNFTPTSGRGNVTFSLTPGPNRYRPGYNNFNTPSLQBSVKATQIRLHFHQ 180  
 DB 430 PDSVNCQLQSNFTPTSGRGNVTFSLTPGPNRYRPGYNNFNTPSLQBSVKATQIRLHFHQ 489  
 QY 181 YTTTAVNLRRHYAVDEIT 200  
 DB 490 YYPAGTVDWRHQQYAVDEIT 509

## RESULT 4

Q8K3K1 ID Q8K3K1 PRELIMINARY; PRT; 1512 AA.  
 AC Q8K3K1  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Usherin.  
 GN USH2A.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22150873; PubMed=12160733;  
 RA Huang D., Eudy J.D., Uzvolgyi E., Davis J.R., Talmadge C.B.,  
 RA Pretto D., Weston M.D., Lehman J.E., Zhou M., Seemayer T.A., Ahmad I.,  
 RA Kimberling W.J., Sumegi J.;  
 RT Identification of the Mouse and Rat Orthologs of the Gene Mutated in  
 RT Usher Syndrome Type IIA and the Cellular Source of USH2A mRNA in  
 RT Retina, a Target Tissue of the Disease.;  
 RL Genomics 80:195-203(2002).  
 DR EMBL; AY077844; RAL78289.1;  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR003961; FN III.  
 DR InterPro; IPR006558; LamG-like.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR001896; LamNT.  
 DR Pfam; PF00041; fn3; 4.  
 DR PRINTS; PR00011; EGF\_LAMININ.  
 DR SMART; SM00180; EGF\_Lam; 10.  
 DR SMART; SM00060; FN3; 4.  
 DR SMART; SM00560; LamGL; 1.  
 DR SMART; SM00136; LamNT; 1.  
 DR PROSITE; PS00022; EGF\_1; 6.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 6.  
 DR Laminin EGF-like domain.  
 KW Laminin EGF-like domain.  
 SQ SEQUENCE 1512 AA; 167591 MW; E37D79F5D00044C1 CRC64;

Query Match 66.7%; Score 728; DB 11; Length 1512;  
 Best Local Similarity 66.7%; Pred. No. 6e-62;  
 Matches 134; Conservative 22; Mismatches 45; Indels 0; Gaps 0;

QY 1 PLAQRVCIPNDAGTADNRVSRNLNPEAHPLSFVNDNDVGTWSVSNVFTNTQLNQGVITIS 60  
 DB 309 PSVQVCIPNGAGDTPEHRMSRLNPEAHPLSFVNDNDVGTWSVSNVFTNTQLNQGVITIS 368  
 QY 61 VDLENGQYQVFIIOFFSPQTEIRIQKENSLEDWEDWQYFARNCGAFGKMGNGDLEK 120  
 DB 369 IDLENGQYQVFIIOFFSPQTEIRIQKENSLEDWEDWQYFARNCGAFGKMGNGDLEK 428

QY 121 PDSVNCQLQSNFTPTSGRGNVTFSLTPGPNRYRPGYNNFNTPSLQBSVKATQIRLHFHQ 180  
 DB 429 PDSVNCQLQSNFTPTSGRGNVTFSLTPGPNRYRPGYNNFNTPSLQBSVKATQIRLHFHQ 488  
 QY 181 YTTTAVNLRRHYAVDEIT 201  
 DB 489 YYPAGTVDWRHQQYAVDEIT 509

## RESULT 5

Q9D1Z8 ID Q9D1Z8 PRELIMINARY; PRT; 214 AA.  
 AC Q9D1Z8  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE A930011D15Rik protein.  
 GN USH2A OR A930011D15Rik.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Retina;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK020845; BAB32226.1; --  
 DR MGD; MGI:1341292; Ush2a.  
 SQ SEQUENCE 214 AA; 24016 MW; C0040881B1DE5702 CRC64;

Query Match 26.9%; Score 294; DB 11; Length 214;  
 Best Local Similarity 74.3%; Pred. No. 1.1e-20;  
 Matches 52; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

QY 1 PLAQRVCIPNDAGTADNRVSRNLNPEAHPLSFVNDNDVGTWSVSNVFTNTQLNQGVITIS 60  
 DB 129 PSVQVCIPNGAGDTPEHRMSRLNPEAHPLSFVNDNDVGTWSVSNVFTNTQLNQGVITIS 188  
 QY 61 VDLENGQYQV 70  
 DB 189 IDLENGQYQV 198

## RESULT 6

O57339 ID O57339 PRELIMINARY; PRT; 569 AA.  
 AC O57339  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Netrin-1 (Fragment).  
 GN NETRIN-1.

OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodidae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA de la Torre J.R., Hoepker V.H., Ming G.-L., Poo M.-m.,  
 RA Tessier-Lavigne M., Hemmati-Brivanlou A., Holt C.E.;  
 RT "Turning of retinal growth cones in a netrin-1 gradient mediated by  
 RT the netrin receptor DCC.";  
 RL Neuron 0:0-0(1997).  
 DR EMBL; AF033341; AAB87983.1; --  
 DR HSP; P02468; 1TLE.  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR001886; LamNT.  
 DR Pfam; PF00053; laminin\_EGF; 3.  
 DR Pfam; PF00055; laminin\_Nterm; 1.  
 DR Pfam; PF01759; NTR; 1.  
 DR PRINTS; PR00011; EGFLAMININ.  
 DR ProDom; PD002082; LamNT; 1.  
 DR SMART; SM00643; C345C; 1.  
 DR SMART; SM00180; EGF Lam; 3.  
 DR SMART; SM00136; LamNT; 1.  
 DR PROSITE; PS00022; EGF 1; 2.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 3.  
 KW Laminin EGF-like domain.  
 FT NON TER 1  
 FT 569 569  
 SQ SEQUENCE 569 AA; 64479 MW; 67B1062E96E8A90 CRC64;  
 Query Match 17.0%; Score 186; DB 13; Length 569;  
 Best Local Similarity 27.4%; Pred. No. 1.4e-09;  
 Matches 58; Conservative 38; Mismatches 92; Indels 24; Gaps 9;  
 QY 5 RYCIPNDAGTADNRVRLN----PEAHPPLSFVND-NDVG--TSWVSNVFTNITQLNQV 57  
 DB 41 RYCVVTEKEDGRFRNCHCNMSDAKRAHPPLSLTLNPNHNLTCWQSE---NYIQYQNV 97  
 QY 58 TISVDLENQYQVFIILIOFPSPQTEIRIQKKNLSLDWEDWQYFARNC-GAPGKNG 116  
 DB 98 TLTLSL-GKKPEVTYVLSQFCSPRESMAIPKSMYDGKSWVPFYSTQCRKYNKPKNA 156  
 QY 117 DIEKPSVNLQLSNFT---PVSRCNVTFSLTPGNVPRGYNVFNFTPSLOESVKATQI 173  
 DB 157 IITKQNEQEAICTDSHTDMHPLSGGLIAFSTL---DGRPSAHDNDSPVLQDWVTATDI 212  
 QY 174 ----RPFHFGQYTTTAVNLHRHYAYDEI 200  
 DB 213 KVAPSLRHTFGDENEDDSLSARDSYFYAVS 244  
 RESULT 7  
 ID O42140 PRELIMINARY; PRT; 603 AA.  
 AC O42140;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Netrin-1a.  
 GN NTN1A.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97454757; PubMed=9268507;  
 RA Lauderdale J.D., Davis N.M., Kuwada J.Y.;  
 RT "Axon tracts correlate with netrin-1 expression in the zebrafish

embryo.";  
 RT Mol. Cell. Neurosci. 9:293-313(1997).  
 DR EMBL; AF002717; AAC60252.1; --  
 DR HSP; P02468; 1KLO.  
 DR ZFIN; ZDB-GENE-990415-169; ntnla.  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR001886; LamNT.  
 DR InterPro; IPR001134; Netrin\_C.  
 DR Pfam; PF00053; laminin\_EGF; 3.  
 DR Pfam; PF00055; laminin\_Nterm; 1.  
 DR Pfam; PF01759; NTR; 1.  
 DR PRINTS; PR00011; EGFLAMININ.  
 DR ProDom; PD002082; LamNT; 1.  
 DR SMART; SM00643; C345C; 1.  
 DR SMART; SM00180; EGF Lam; 3.  
 DR SMART; SM00136; LamNT; 1.  
 DR PROSITE; PS00022; EGF 1; 2.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 3.  
 KW Laminin EGF-like domain.  
 SQ SEQUENCE 603 AA; 67681 MW; 82EDB32E3E383F37 CRC64;  
 Query Match 16.9%; Score 185; DB 13; Length 603;  
 Best Local Similarity 27.0%; Pred. No. 1.9e-09;  
 Matches 58; Conservative 41; Mismatches 86; Indels 30; Gaps 10;  
 QY 5 RYCIPNDAGD-----TADNRVSLNPEAHPPLSFVND-NDVG--TSWVSNVFTNITQLN 54  
 DB 75 RYCVVTEKEDGRHRNCHTCDASDPKKN---HPPAYLTDLNPNHNLTCWQSD---NVLQVP 128  
 QY 55 QGVTSVDLENQYQVFIILIOFPSPQTEIRIQKKNLSLDWEDWQYFARNC-GAPGK 113  
 DB 129 QNVTLTSL-GKKPEVTYVLSQFCSPRESMAIPKSMYDGKSWVPFYSTQCRKYNK 187  
 QY 114 NGDLEKPSVNLQLSNFT---PVSRCNVTFSLTPGNVPRGYNVFNFTPSLOESVK 170  
 DB 198 SKAITKQNEQEAICTDSHTDMHPLSGGLIAFSTL---DGRPSAHDNDSPVLQDW 243  
 QY 171 TOI-----RPFHFGQYTTTAVNLHRHYAYDEI 200  
 DB 244 TDIKVTFSLRHTFGDENEDDSLSARDSYFYAVS 278  
 RESULT 8  
 ID Q92429 PRELIMINARY; PRT; 604 AA.  
 AC Q92429;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Netrin-1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21267048; PubMed=11356879;  
 RA Manitt C., Colicos M.A., Thompson K.M., Rousselle E., Peterson A.C.,  
 RA Kennedy T.E.;  
 RT "Widespread Expression of Netrin-1 by Neurons and Oligodendrocytes in  
 RT the Adult Mammalian Spinal Cord.";  
 RL J. Neurosci. 21:3911-3922(2001).  
 DR EMBL; AY028417; AAK17014.1; --  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR001886; LamNT.  
 DR InterPro; IPR001134; Netrin\_C.  
 DR Pfam; PF00053; laminin\_EGF; 3.  
 DR Pfam; PF01759; NTR; 1.  
 DR PRINTS; PR00011; EGFLAMININ.  
 DR ProDom; PD002082; LamNT; 1.

[illegible]





```

Query Match      14.7%; Score 160; DB 4; Length 529;
Best Local Similarity 26.2%; Pred. No. 4.4e-07;
Matches 58; Conservative 34; Mismatches 83; Indels 46; Gaps 11;

QY 3 AQRVCIPNDAGDTADNRVSRNLNPEA-HPLSFVND---NDVGTSMVSNVFTNITQLNQGYT 58
Db 70 AGAHCORCDAAD-----PORHNASYLTDFHSQDESTWQSPSWAFGVQYPTSVN 119
QY 59 ISVDLENGQYQVYIIIOFFSPOPTTEIRIQKKNESLWEDWQYFARNCGAFGMKRNQDGL 118
Db 120 ITRLGKA-YEITYVRLKFHTSRPESFALYKRSRADGPWEYQFYASAC-----QKTYG-- 172
QY 119 EKPDSVNCI-----OLSNPTPYSEGNVTFSLTPGPNYRPGYNNFYNTPSIQ 165
Db 173 -RPEG-QCLRPGEDERVAFTSEFSDISPLSGGNVAFSTL-----EGRPSAYNFESPGIQ 226
QY 166 ESKVATOI-----RFHFHGQYTTTAVNLRHRYAVDEIT 201
Db 227 EWTSTELLISLDRLNTGDDIFKPKV-LQSYYYAVSDFS 266

```

Search completed: December 31, 2003, 12:03:52  
Job time : 12.9842 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2003, 11:57:49 ; Search time 1.2569 Seconds  
(without alignments)  
4722.097 Million cell updates/sec

Title: US-09-970-318-1

Perfect score: 126

Sequence: 1 QAPPTQGPPTVWKISPTSLRIE 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archesp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	126	100.0	1461	11 Q9JLP3	Q9JLP3 mus musculus
2	108	85.7	1512	11 Q8K3K1	Q8K3K1 rattus norv
3	71	56.3	1546	4 Q75445	Q75445 homo sapien
4	71	56.3	1546	4 Q9NS27	Q9NS27 homo sapien
5	54.5	43.3	4340	2 Q30764	Q30764 streptomyce
6	53	42.1	434	16 Q8X5J0	Q8X5J0 escherichia
7	52.5	41.7	1286	5 Q9TXR6	Q9TXR6 caenorhabdi
8	51	40.5	1243	12 Q8JFK6	Q8JFK6 heliothis z
9	50	39.7	647	3 Q8WZL8	Q8WZL8 varrovia li
10	50	39.7	946	12 Q69139	Q69139 human herpe
11	50	39.7	981	11 Q8CBQ0	Q8CBQ0 mus musculu
12	50	39.7	981	11 Q8C881	Q8C881 mus musculu
13	50	39.7	989	11 Q99ML9	Q99ML9 mus musculu
14	50	39.7	1133	3 Q8NIT4	Q8NIT4 neurospora
15	49.5	39.3	564	11 Q8K4M2	Q8K4M2 rattus norv
16	49.5	39.3	570	11 Q99J80	Q99J80 rattus norv

17	49	38.9	286	16 Q8YXG6	Q8YXG6 anabaena sp
18	49	38.9	318	4 Q8NFG7	Q8NFG7 homo sapien
19	49	38.9	321	11 Q9QXFS	Q9QXFS mesocricetu
20	49	38.9	328	4 Q8N6A6	Q8N6A6 homo sapien
21	49	38.9	328	4 Q8N6A7	Q8N6A7 homo sapien
22	49	38.9	332	6 Q9N257	Q9N257 oryctolagus
23	49	38.9	335	4 Q96I06	Q96I06 homo sapien
24	49	38.9	336	4 Q8NFD3	Q8NFD3 homo sapien
25	49	38.9	356	4 Q9HAA9	Q9HAA9 homo sapien
26	49	38.9	364	10 Q8S0C5	Q8S0C5 oryza sativ
27	49	38.9	394	17 Q9HLM3	Q9HLM3 thermoplas
28	49	38.9	438	16 Q9K4A7	Q9K4A7 streptomyce
29	49	38.9	1135	13 Q9YHD3	Q9YHD3 xenopus lae
30	49	38.9	3164	12 Q69088	Q69088 human herpe
31	48.5	38.5	521	4 Q8TEN0	Q8TEN0 homo sapien
32	48.5	38.5	778	4 Q8WY92	Q8WY92 homo sapien
33	48.5	38.5	834	4 Q81Z84	Q81Z84 homo sapien
34	48	38.1	216	11 Q9CWE8	Q9CWE8 mus musculu
35	48	38.1	290	12 Q9Q8U2	Q9Q8U2 shope fibro
36	48	38.1	333	11 Q91Y74	Q91Y74 mus musculu
37	48	38.1	333	11 Q92IR5	Q92IR5 mus musculu
38	48	38.1	333	11 P97354	P97354 m cmp-n-ace
39	48	38.1	370	10 Q64895	Q64895 arabidopsis
40	48	38.1	396	10 Q9SJF3	Q9SJF3 arabidopsis
41	48	38.1	421	5 Q9VM68	Q9VM68 drosophila
42	48	38.1	487	16 Q8XV67	Q8XV67 raistonia s
43	48	38.1	652	3 Q9Y835	Q9Y835 penicillium
44	48	38.1	838	5 Q9VQA9	Q9VQA9 drosophila
45	48	38.1	2338	5 Q94269	Q94269 caenorhabdi

## ALIGNMENTS

RESULT 1  
Q9JLP3 PRELIMINARY; PRT; 1461 AA.  
ID Q9JLP3  
AC Q9JLP3  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Putative extracellular matrix protein MUSH2A.  
GN MUSH2A.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rudy J.D., Talmadge C.B., Weston M.D., Yao S.-F., Cosgrove D.,  
RA Ahmad I., Kimberling W., Sumegi J.;  
RT "Isolation and Characterization of the Murine Homolog of the Usher  
RT Syndrome Type 2A Gene."  
RL Submitted (MAY-1999) to the ENBL/GenBank/DBJ databases.  
DR EMBL; AF151717; RAP70550.1; -;  
DR HSSP; P02468; TITLE.  
DR MGD; MGI:1341292; Ush2a.  
DR InterPro; IPR006209; EGF like.  
DR InterPro; IPR003961; FN III.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR InterPro; IPR001886; Laminin\_EGF.  
DR Pfam; PF00041; fn3; 4.  
DR Pfam; PF00053; laminin\_Nterm; 1.  
DR Pfam; PF00055; laminin\_Nterm; 1.  
DR PRINTS; PR00011; EGF\_LAMININ.  
DR SMART; SM00180; EGF\_Lam; 8.  
DR SMART; SM00060; FN3; 3.  
DR SMART; SM00136; Laminin; 1.  
DR PROSITE; PS00022; EGF 1; 7.  
DR PROSITE; PS01248; LAMININ TYPE EGF; 7.  
KW Laminin EGF-like domain; Matrix protein.  
SQ SEQUENCE 1461 AA; 160989 MW; 114436D877C38A15 CRC64;

```

Query Match      100.0%; Score 126; DB 11; Length 1461;
Best Local Similarity 100.0%; Pred. No. 4.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QAPPQTQGPPTWKISPTLRIE 23
    |||||
DB 1237 QAPPQTQGPPTWKISPTLRIE 1259
    |||||

RESULT 2
Q8K3K1 PRELIMINARY; PRT; 1512 AA.
AC Q8K3K1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Usherin.
GN USH2A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22150873; PubMed=12160733;
RA Huang D., Eudy J.D., Uzvolgyi E., Davis J.R., Talmadge C.B.,
RA Pretto D., Weston M.D., Lehman J.E., Zhou M., Seemayer T.A., Ahmad I.,
RA Kimberling W.J., Sumegi J.;
RT "Identification of the Mouse and Rat Orthologs of the Gene Mutated in
RT Usher Syndrome Type IIA and the Cellular Source of USH2A mRNA in
RT Retina, a Target Tissue of the Disease.";
RL Genomics 80:195-203(2002).
DR EMBL; AY077844; AAL78289.1; -.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR006558; LamG like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001886; LamNT.
DR Pfam; PF00041; fn3; 4.
DR PRINTS; PR00011; EGF_Lam; 10.
DR SMART; SM00560; FN3; 4.
DR SMART; SM00060; LamGL; 1.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 6.
KW Laminin EGF-like domain.
SQ SEQUENCE 1512 AA; 167591 MW; E37D79F5D00044C1 CRC64;

Query Match      85.7%; Score 108; DB 11; Length 1512;
Best Local Similarity 82.6%; Pred. No. 2.1e-07;
Matches 19; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 QAPPQTQGPPTWKISPTLRIE 23
    |||||
DB 1236 QAPPQTQGPPTWKISPTLRIE 1258
    |||||

RESULT 3
Q75445 PRELIMINARY; PRT; 1546 AA.
AC Q75445;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Usher syndrome type IIA protein.
GN USH2A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=98288356; PubMed=9624053;
RA Eudy J.D., Weston M.D., Yao S.F., Hoover D.M., Rehm H.L., Ahmad I.,
RA Ma-Edmonds M., Yan D., Cheng J.J., Beisel K.W., Ayuso C., Cremers C.,
RA Davenport S., Moller C., Talmadge C.B., Tamayo M., Swaroop A.,
RA Morton C.C., Kimberling W.J., Sumegi J.;
RT "Mutation of a gene encoding a protein with extracellular matrix
RT motifs in Usher syndrome type IIA.";
RL Science 280:1753-1757(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Eudy J.D., Yao S.F., Cheng J.J., Weston M.D., Sumegi J.;
RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF055580; AAC23748.2; -.
DR HSSP; P02468; IKLO.
DR Genew; HGNC:12601; USH2A.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR006558; LamG like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001886; LamNT.
DR Pfam; PF00041; fn3; 4.
DR Pfam; PF00053; laminin_Nterm; 1.
DR PRINTS; PR00011; EGF_LAMININ.
DR SMART; SM00180; EGF_Lam; 10.
DR SMART; SM00060; FN3; 4.
DR SMART; SM00560; LamGL; 1.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; 7.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
KW Laminin EGF-like domain.
SQ SEQUENCE 1546 AA; 170985 MW; 49CD0A95A614959F CRC64;

Query Match      56.3%; Score 71; DB 4; Length 1546;
Best Local Similarity 60.9%; Pred. No. 0.072;
Matches 14; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 QAPPQTQGPPTWKISPTLRIE 23
    |||||
DB 1240 QAPPQTQGPPTWKISPTLRIE 1262
    |||||

RESULT 4
Q9NS27 PRELIMINARY; PRT; 1546 AA.
AC Q9NS27;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Usher syndrome type IIA protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20206315; PubMed=10729113;
RA Weston M.D., Eudy J.D., Fugita S., Yao S.-F., Usami S., Cremers C.,
RA Greenburg J., Ramesar R., Martini A., Moller C., Smith R.J.,
RA Sumegi J., Kimberling W.J.;
RT "Genomic structure and identification of novel mutations in usherin,
RT the gene responsible for Usher syndrome type IIA.";
RL Am. J. Hum. Genet. 66:1199-1210(2000).
DR EMBL; AF091889; AAF75819.1; -.
DR EMBL; AF091873; AAF75819.1; JOINED.
DR EMBL; AF091874; AAF75819.1; JOINED.
DR EMBL; AF091875; AAF75819.1; JOINED.
DR EMBL; AF091876; AAF75819.1; JOINED.
DR EMBL; AF091877; AAF75819.1; JOINED.
DR EMBL; AF091878; AAF75819.1; JOINED.
DR EMBL; AF091879; AAF75819.1; JOINED.
DR EMBL; AF091880; AAF75819.1; JOINED.
DR EMBL; AF091881; AAF75819.1; JOINED.

```



```

DE M01E10.2 protein.
GN M01E10.2
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Geisel C., Kramer J., Smith A.;
RT "The sequence of C. elegans Cosmid M01E10.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF099921; AAC68807.1; -.
DR WormRep; M01E10.2; CE19534.
DR InterPro; IPR001507; Endoglin/CD105.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR Pfam; PF00100; zona_pellucida; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00327; VWA; 1.
DR SMART; SM00241; ZP; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 1286 AA; 140794 MW; 95781989BC127CCE CRC64;

Query Match 41.7%; Score 52.5; DB 5; Length 1386;
Best Local Similarity 50.0%; Pred. No. 35;
Matches 11; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 1 QAPPOTGPPPTVWKISPTLRI 22
:|||||:
DB 336 EAPPTTQTPTTSW-TRPTDGRV 356

RESULT 8
Q8JJK6 PRELIMINARY; PRT; 1243 AA.
AC Q8JJK6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Macin-like glycoprotein 900.
GN ORP107.
OS Heliothis zea virus 1.
OC Viruses; unclassified viruses.
OX NCBI_TaxID=29250;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22174892; PubMed=12186886;
RA Cheng C.H., Liu S.M., Chow T.Y., Hsiao Y.Y., Wang D.P., Huang J.J.,
RA Chen H.H.;
RT "Analysis of the Complete Genome Sequence of the Hs-1 Virus Suggests
RT that It Is Related to Members of the Baculoviridae.";
RL J. Virol. 76:9024-9034 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Cheng C.H., Liu H.M., Hsiao Y.Y., Chow T.Y., Chen H.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]

```

```

RP SEQUENCE FROM N.A.
RA Cheng C.H., Hsiao Y.Y., Liu S.M., Chow T.Y., Chen H.H.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF451898; AAN04401.1; -.
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 1243 AA; 136853 MW; B5400BD894B5CBA9 CRC64;

Query Match 40.5%; Score 51; DB 12; Length 1243;
Best Local Similarity 47.8%; Pred. No. 56;
Matches 11; Conservative 4; Mismatches 4; Indels 4; Gaps 1;

QY 3 PPOTGPPPTV----WKISPTLRI 21
|||||:
DB 830 PELTQPPTTINLTTEKLSPTLK 852

RESULT 9
Q8WZL8 PRELIMINARY; PRT; 647 AA.
AC Q8WZL8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Transcriptional repressor, TUP1.
GN Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=POLA;
RA Cabello J., Dominguez A.;
RT "The Yarrowia lipolytica TUP1, a gene involved in the yeast-to-hypha
RT transition.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=POLA;
RA Cabello J.;
RT "New strategies for studying dimorphism. Isolation and characterization
RT of a transcriptional repressor, YITUP1, from Yarrowia lipolytica.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ252128; CAC81004.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 6.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS50082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 647 AA; 71292 MW; 7A7CD9C988A35DC5 CRC64;

Query Match 39.7%; Score 50; DB 3; Length 647;
Best Local Similarity 50.0%; Pred. No. 40;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 QAPPOTGPPPTVWKISPTLRI 20
|||||:
DB 279 QGPPQQQQPPQQQQSQSPEM 298

RESULT 10
Q69139 PRELIMINARY; PRT; 946 AA.
AC Q69139;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nuclear antigen EBNA-3B.

```

OS Human herpesvirus 4 (Epstein-Barr virus).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Lymphocryptovirus.  
 OX NCBI\_TaxID=10376;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90347798; PubMed=2166806;  
 RA Sample J., Young L., Martin B., Chatman T., Kieff B.D, Rickinson A.;  
 RT "Epstein-Barr virus types 1 and 2 differ in their EBNA-3A, EBNA-3B and  
 RT EBNA-3C genes.";  
 RL J. Virol. 64:4084-4092(1990).  
 DR EMBL; M34440; AAA45894.1; -;  
 SQ SEQUENCE 946 AA; 104103 MW; 7F428D409134E869 CRC64;

Query Match 39.7%; Score 50; DB 12; Length 946;  
 Best Local Similarity 52.9%; Pred. No. 60;  
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 PQQTGGPPTVWKISPTL 20  
 DB 753 PPQGGPPTAMQLSPAL 769

RESULT 11  
 Q8C8C0 ID Q8C8C0 PRELIMINARY; PRT; 981 AA.  
 AC Q8C8C0;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Arkadia.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK036351; BAC29394.1; -;  
 SQ SEQUENCE 981 AA; 106917 MW; 14368A4582C1F5D5 CRC64;

Query Match 39.7%; Score 50; DB 11; Length 981;  
 Best Local Similarity 62.5%; Pred. No. 62;  
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 APPQTGGPPTVWKISP 17  
 DB 666 APPQTGGPPTVWKISP 681

RESULT 12  
 Q8C881 ID Q8C881 PRELIMINARY; PRT; 981 AA.  
 AC Q8C881;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Arkadia.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK048110; BAC33245.1; -;  
 SQ SEQUENCE 981 AA; 106928 MW; E421C32F1A64CFEE CRC64;  
 Query Match 39.7%; Score 50; DB 11; Length 981;  
 Best Local Similarity 62.5%; Pred. No. 62;  
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 APPQTGGPPTVWKISP 17  
 DB 666 APPQTGGPPTVWKISP 681

RESULT 13  
 Q99ML9 ID Q99ML9 PRELIMINARY; PRT; 989 AA.  
 AC Q99ML9;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Arkadia.  
 GN ARK OR ARKADIA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/Sv;  
 RX MEDLINE=21195971; PubMed=11298452;  
 RA Episkopou V., Arkell R., Timmons P.M., Walsh J.J., Andrew R.L.,  
 RA Swan D.;  
 RT "Induction of the mammalian node requires Arkadia function in the  
 RT extraembryonic lineages.";  
 RL Nature 410:825-830(2001).  
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 DR EMBL; AF330197; AAK38272.1; -;  
 DR MGD; MGI:1934919; Ark.  
 DR InterPro; IPR001211; PhospholipaseA2.  
 DR Pfam; PF00097; Zf-C3HC4\_1.  
 DR SMART; SM00184; RING; 1.  
 DR PROSITE; PS00118; PA2\_HIS; 1.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KW Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 989 AA; 107896 MW; 212E3C37BC70DCB5 CRC64;

Query Match 39.7%; Score 50; DB 11; Length 989;  
 Best Local Similarity 62.5%; Pred. No. 62;  
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 APPQTGGPPTVWKISP 17  
 DB 666 APPQTGGPPTVWKISP 681

RESULT 14  
 Q8NIT4 ID Q8NIT4 PRELIMINARY; PRT; 1133 AA.  
 AC Q8NIT4;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN B13H18.270.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]

```

RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RN Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL807374; CAD37073.1; -
KW Hypothetical protein.
SQ SEQUENCE 1133 AA; 124593 MW; 7B0A9C52CF59DF17 CRC64;

Query Match          39.7%; Score 50; DB 3; Length 1133;
Best Local Similarity 50.0%; Pred. No. 72;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 PPOTQGPPTVWKISPT 18
DB 446 PTPQSPPTLYRHSPS 461

RESULT 15
Q8K4M2
ID Q8K4M2 PRELIMINARY; PRT; 564 AA.
AC Q8K4M2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE FAD-dependent sulphydryl oxidase.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Benayoun B., Ennard-Reve A., Ennard P.;
RT "Rat FAD-dependent sulphydryl oxidase (SOX) mRNA.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF217799; AAM67412.1; -
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR006662; Thiored.
DR InterPro; IPR006663; Thioredox_dom2.
DR Pfam; PF00085; Thiored; 1.
DR PRINTS; PR00421; THIOREDOXIN.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
KW Redox-active center.
SQ SEQUENCE 564 AA; 62856 MW; 33B0A60AE6AF1A84 CRC64;

Query Match          39.3%; Score 49.5; DB 11; Length 564;
Best Local Similarity 43.5%; Pred. No. 41;
Matches 10; Conservative 2; Mismatches 4; Indels 7; Gaps 1;

QY 1 QAPPOTQGP-----PTVWKIS 16
DB 269 EAPPTTAAPVTPDKIAPTIVKFA 291

```

Search completed: December 31, 2003, 12:03:50  
Job time : 2.2569 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 06:39:50 ; Search time 14799 Seconds  
(without alignments)  
17498.340 Million cell updates/sec

Title: US-09-970-318-3  
Perfect score: 6330  
Sequence: 1 tggctgctgcagaataact.....gcaaaaaaaaaaaaaaaa 6330

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em.sy.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	6307.6	99.6	6332	9	AF055580	AF055580 Homo sapi
2	2748	43.4	5126	10	AF151717	AF151717 Mus muscu
3	2683.6	42.4	6146	10	AY077844	AY077844 Rattus no
4	1546.2	24.4	2652	9	HSUSH2A17	AF091889 Homo sapi
5	1546.2	24.4	115940	9	AC119429	AC119429 Homo sapi
6	1546.2	24.4	163815	9	AC138024	AC138024 Homo sapi
7	1531	24.2	184849	2	AC021209	AC021209 Homo sapi
8	683.4	10.8	2749	9	HSUSH2A01	AF091873 Homo sapi
9	645.8	10.2	1593	9	HSUSH2A10	AL445650 Human DNA
10	645.8	10.2	184849	2	AC021209	AF091882 Homo sapi
11	645.8	10.2	290708	2	AL161841	AC021209 Homo sapi
12	496.4	7.9	2840	9	HSUSH2A14	AL161641 Homo sapi
13	476.4	7.5	2749	6	AX347325	AF091886 Homo sapi
14	437	6.9	2749	6	AX347324	AX347325 Sequence
15	437	6.9	2749	6	AX347324	AX347324 Sequence
16	432.4	6.8	152888	10	AC121892	AC121892 Mus muscu
17	427.6	6.8	241363	2	AC106298	AC106298 Rattus no
18	412.8	6.5	290708	2	AL161841	AL161641 Homo sapi
19	301	4.8	3175	9	HSUSH2A05	AF091877 Homo sapi
20	301	4.8	93856	9	AL358858	AL358858 Human DNA
21	257.6	4.1	257291	2	AC122122	AC122122 Mus muscu
22	250	3.9	317136	2	AC098150	AC098150 Rattus no
23	233.8	3.7	290002	2	AC098145	AC098145 Rattus no
24	204	3.2	317136	2	AC098150	AC098150 Rattus no
25	199.2	3.1	1348	9	HSUSH2A07	AC098145 Rattus no
26	199	3.1	883	9	HSUSH2A09	AF091879 Homo sapi
27	185.2	2.9	1047	9	HSUSH2A11	AF091881 Homo sapi
28	178.6	2.8	106863	5	AL928842	AL928842 Zebrafish
29	176	2.8	1255	9	HSUSH2A02	AF091883 Homo sapi
30	174.8	2.8	1329	9	HSUSH2A15	AF091874 Homo sapi
31	166.4	2.6	1268	9	HSUSH2A12	AF091887 Homo sapi
32	162.4	2.6	1364	9	HSUSH2A13	AF091884 Homo sapi
33	162.2	2.6	269897	2	AC121799	AF091885 Homo sapi
34	155.8	2.5	726	11	BV028598	AC121799 Mus muscu
35	147	2.3	1662	9	HSUSH2A16	BV028598 S212P6051
36	136.8	2.2	948	9	HSUSH2A03	AF091888 Homo sapi
37	134.8	2.1	1243	9	HSUSH2A08	AF091875 Homo sapi
38	117.4	1.9	5433	6	AX045107	AF091880 Homo sapi
39	117.4	1.9	5433	6	AX045324	AX045107 Sequence
40	117.4	1.9	5433	6	AX463744	AX045324 Sequence
41	117.4	1.9	5613	6	AX045105	AX463744 Sequence
42	117.4	1.9	5613	6	AX045322	AX045105 Sequence
43	117.4	1.9	5613	6	AX395746	AX045322 Sequence
44	117.4	1.9	5613	6	AX409725	AX395746 Sequence
45	117.4	1.9	5613	6	AX463742	AX409725 Sequence
						AX463742 Sequence

# ALIGNMENTS

RESULT 1  
AF055580  
LOCUS AF055580 6332 bp mRNA linear PRI 21-MAR-2002  
DEFINITION Homo sapiens Usher syndrome type IIA protein (USH2A) mRNA, complete cds.  
ACCESSION AF055580  
VERSION AF055580.2 GI:19584001  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 6332)  
AUTHORS Eudy,J.D., Weston,M.D., Yao,S.F., Hoover,D.M., Rehm,H.L., Ahmad,I.,  
Ma-Edmonds,M., Yan,D., Cheng,J.J., Beisel,K.W., Ayuso,C.,

Creemers, C., Davenport, S., Moller, C., Talmadge, C.B., Tamayo, M., Swaroop, A., Morton, C.C., Kimberling, W.J., and Sumegi, J.  
Mutation of a gene encoding a protein with extracellular matrix motifs in Usher syndrome type IIa  
Science 280 (5370), 1753-1757 (1998)

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REMARK

## COMMENT

## FEATURES

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

## source

Creemers, C., Davenport, S., Moller, C., Talmadge, C.B., Tamayo, M., Swaroop, A., Morton, C.C., Kimberling, W.J., and Sumegi, J.  
Mutation of a gene encoding a protein with extracellular matrix motifs in Usher syndrome type IIa  
Science 280 (5370), 1753-1757 (1998)

98288336

9624053

2 (bases 1 to 6332)

Eudy, J.D., Yao, S.F., Cheng, J.J., Weston, M.D. and Sumegi, J.

Direct Submission

Submitted (24-MAR-1998) Pathology/Microbiology, University of

Nebraska-Medical Center, 600 South 42nd Street, Omaha, NE

68198-5660, USA

3 (bases 1 to 6332)

Eudy, J.D., Yao, S.F., Cheng, J.J., Weston, M.D. and Sumegi, J.

Direct Submission

Submitted (21-MAR-2002) Pathology/Microbiology, University of

Nebraska-Medical Center, 600 South 42nd Street, Omaha, NE

68198-5660, USA

Sequence update by submitter

On Mar 21, 2002 this sequence version replaced gi:3241997.

Location/Qualifiers

1..6332

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/chromosome="1"

/map="1q41; between AFM2682D1 and AFM144XF2"

1..6332

/gene="USH2A"

388..5028

/gene="USH2A"

/note="putative extracellular matrix component or cell

adhesion molecule"

/codon\_start=1

/product="Usher syndrome type IIa protein"

/protein\_id="AAC23748.2"

/db\_xref="GI:19584002"

/translation="MNCPLSLGSGFLFQVLEMLFAPASISLTERGLPRLNNGV

AFKYSIVPTQVACGLPRSTFCSSAAASIQCTQRCPLQDCPSSSHYTYALFS

AGLSCITPDNDLHNAHSASPIFNHNSCFSSPPSKLMASFTLAWLKPQQV

NVCLEKTVQGIQVFLKTSKEMFYRTVNGLPPIKMTLGRILVKKWHLISVQ

HQTKLSFPIGVEKDETPFNARTLSGISTDFASGVTOIGSLNGLEQFVGRMODFLY

QVALNRIILVFSGLDLAQAHCRCPSGSHPRVPLAORYCIPNDAGDPAIDNRVSR

LNPEAHLPSFYNDNDVGSVSNVFTNITQLNGVTLSVDLENGOYQVYIIIOFSP

QPTERIQRKNSLDHEDWQYFARNCGAFCKMKGDLKEDSVNCLQLSNFTFYSRG

NVTSILTPGPNRYNNFYNTPSLOESVRATQIRHFHGGYYTTTAVNLRRHYA

VDEITISRCQCHHADNCDTTSQPYRCLCSQESFTEGLHCDRLPLNDKPPRQGY

VYAFNCKPCQNSHKSCHYNI SVDPFPFHFPRGGGVCDCEHNTTGRNCBLCKQYF

PROVAGPSAIDVCKPCDDTVTRNGSILCDOIGGCNCKRHSVGRQCNQONGFYN

LOELDPDSCNCTSVTDGDTCHONSQCKKANVIGLRCHDCHNCPKFLRSFN

DVGCPCQNLGHSNKCNPHSQGCCKEAKGLQCDTCENFYGLDVTNCKACDC

TAGSLPFTVCAKATQCCICKPNVEGRQCNKLEGNFYLRQNSFLCLPNCNCKDTGTIN

GSLLCNKSGQCPCKGLVLRQNCQEPHYRLNLTIDNPHQMCCECSLGLPTGICD

PLSGQCLVPRQKRCNQCPQFYSIPGNATGLPCSTHTTGAHVNI CNSLTQCVG

QDASLAGORCDQCKHDFGFPOTGRCPQCNCHLSGALNETCHLVTCQCPKQPVTS

KDCAVPSASHLDVNNLIGSKTPEQPPRQVQSSANLNSWSPDSNAHLATYS

LRDGFPIYTHEDQYPSIQFLDTLLPYTKYSYIETNVHSGTSASVATYKRG

VPEGLMYSIIIPQSDSVTLTTLNSGPIBKRYILSLCAPLAGGQPCVSYGHEYS

ATIMNLVFPKDYFQVQATSGGLSLPITVTVAQPPRLSPKQKISSTELHYE

WSPPAELNGIIIRYELNMRRLSKETTSBESRVSQSLSPHSFVSEANENALKEP

QTMWITGLEPPTYKVEFVLA VNWAGSVSSAWBERTGESAPVFWIPSPVPLSYSL

NISWKPADNTRGVGVYDINLSESPQOSIPWAPSQILHATAKSOBLSVTVSELKP

YRIYFTILCNVSCVTSASGAGTLLAQAQLRPPLVKGINTTILHLKPPPEELN

GPSPIQLERRSSLPALMTTMMKGFIRFGNGICKFSSHTHPVNDFTGKCV"

1270..3522

/misc\_feature

/gene="USH2A"

/note="encodes laminin epidermal growth factor repeats"

3640..4872

/misc\_feature

/gene="USH2A"

/note="encodes fibronectin type III repeats"

1847 a 1336 g 1854 t 1 others

BASE COUNT

ORIGIN

1018

1020

1022

1024

1026

1028

1030

1032

1034

1036

1038

1040

1042

1044

1046

1048

1050

1052

1054

1056

1058

1060

1062

1064

1066

1068

1070

1072

1074

1076

1078

1080

1082

1084

1086

1088

1090

1092

1094

1096

1098

1100

1102

1104

1106

1108

1110

1112

1114

1116

1118

1120

1122

1124

1126

1128

1130

1132

1134

1136

1138

1140

1142

1144

1146

1148

1150

1152

1154

1156

1158

1160

1162

1164

1166

1168

1170

1172

1174

1176

1178

1180

1182

1184

1186

1188

1190

1192

1194

1196

1198

1200

1202

1204

1206

1208

1210

1212

1214

1216

1218

1220

1222

1224

1226

1228

1230

1232

1234

1236

1238

1240

1242

1244

1246

1248

1250

1252

1254

1256

1258

1260

1262

1264

1266

1268

1270

1272

1274

1276

1278

1280

1282



Db 3181 AATCAGTGTCAACGAGTTTATATATTTCTCCAGGCAATGCCACTGGCTGCCATGC 3240  
Qy 3239 TCATGCCATACAACTGGCGCAGTTAATCAATCTGTAAATAGCCCTGACTGGTCAAGTGTGT 3298  
Db 3241 TCATGCCATACAACTGGCGCAGTTAATCAATCTGTAAATAGCCCTGACTGGTCAAGTGTGT 3300  
Qy 3299 TGCCAAAGATCTTCATTTCTGGCGCAGTTGTGACCAATGCAAGACCAATTTACTTTGGA 3358  
Db 3301 TGCCAAAGATCTTCATTTCTGGCGCAGTTGTGACCAATGCAAGACCAATTTACTTTGGA 3360  
Qy 3359 TTTGATCTCTCAGACTGGAAGATGTCAAGCTTTGTAAATGTCTCTCAGGAGCCCTTCAAT 3418  
Db 3361 TTTGATCTCTCAGACTGGAAGATGTCAAGCTTTGTAAATGTCTCTCAGGAGCCCTTCAAT 3420  
Qy 3419 GAACCTGTCTCAGCTGGTCAAGGCAAGTGTCTGTAAACAAATTTGTCACTGGCTCAAG 3478  
Db 3421 GAACCTGTCTCAGCTGGTCAAGGCAAGTGTCTGTAAACAAATTTGTCACTGGCTCAAG 3480  
Qy 3479 TGTGATCTGTGTCTCCAGTCAAGGCACTTGGATGTCAACATCTATTGGGTGGCAGC 3538  
Db 3481 TGTGATCTGTGTCTCCAGTCAAGGCACTTGGATGTCAACATCTATTGGGTGGCAGC 3540  
Qy 3539 AAAAATCCATTCAGCAACCTCCGCCCAGAGGCAAGTTCAAAGTTCCTCTGCTATCAAT 3598  
Db 3541 AAAAATCCATTCAGCAACCTCCGCCCAGAGGCAAGTTCAAAGTTCCTCTGCTATCAAT 3600  
Qy 3599 CTCTCTGGAGTCACTGTGATCTCCAAATGCCACTGGCTTACTTACAGTTACTCAGG 3658  
Db 3601 CTCTCTGGAGTCACTGTGATCTCCAAATGCCACTGGCTTACTTACAGTTACTCAGG 3660  
Qy 3659 GATGGTTTTGAATCTACAAAGAGGATCAATACCCATACAGTATTCAATTTCTTTA 3718  
Db 3661 GATGGTTTTGAATCTACAAAGAGGATCAATACCCATACAGTATTCAATTTCTTTA 3720  
Qy 3719 GACACAGACTGTGTACCAATATACCAATATTTCTATTACATTCAGACCAACCAATGTGCAT 3778  
Db 3721 GACACAGACTGTGTACCAATATACCAATATTTCTATTACATTCAGACCAACCAATGTGCAT 3780  
Qy 3779 GGTTCACAAAGAGTGTAGCTGTCACTTACAAAGCAAAACAGGGGTCCAGAGGGAAAC 3838  
Db 3781 GGTTCACAAAGAGTGTAGCTGTCACTTACAAAGCAAAACAGGGGTCCAGAGGGAAAC 3840  
Qy 3839 TTGACTTTAGTTATATATCTTCTATTGGCTCAGACTCTGTGACATTTACTTGCAACCA 3898  
Db 3841 TTGACTTTAGTTATATATCTTCTATTGGCTCAGACTCTGTGACATTTACTTGCAACCA 3900  
Qy 3899 CTCTCAAAATCAATCTGTGTCCCATAGAGAAATATATTTGTCTGTGCCCCCTTTGGCTGGT 3958  
Db 3901 CTCTCAAAATCAATCTGTGTCCCATAGAGAAATATATTTGTCTGTGCCCCCTTTGGCTGGT 3960  
Qy 3959 GGTTCAGCCATGTGTCTTCTTCAAGAGTGTATGAAACCTCAGTACCATCTGGAATCTGGTT 4018  
Db 3961 GGTTCAGCCATGTGTCTTCTTCAAGAGTGTATGAAACCTCAGTACCATCTGGAATCTGGTT 4020  
Qy 4019 CCATTTGCCAAGTACGATTTTCTGTACAGCGGTGTACTAGCGGGGCTGTTTACACAGC 4078  
Db 4021 CCATTTGCCAAGTACGATTTTCTGTACAGCGGTGTACTAGCGGGGCTGTTTACACAGC 4080  
Qy 4079 TTGCCCATACAGTACACAGCCAGGCCCTCCCAAGAGCTTAAGTCCACCTAGATG 4138  
Db 4081 TTGCCCATACAGTACACAGCCAGGCCCTCCCAAGAGCTTAAGTCCACCTAGATG 4140  
Qy 4139 CAGAAATCAGTTCTACAGAACTTCATGTAGAAATGGTCTCCACAGCGGAACTAAATGGA 4198  
Db 4141 CAGAAATCAGTTCTACAGAACTTCATGTAGAAATGGTCTCCACAGCGGAACTAAATGGA 4200  
Qy 4199 ATAAATTAAGATATGAACATATACAGAGAACTGAGATCTACTAAGAAACCAATCT 4258  
Db 4201 ATAAATTAAGATATGAACATATACAGAGAACTGAGATCTACTAAGAAACCAATCT 4260  
Qy 4259 GAGAAAGTCCAGTTTTTTCAGACAGTGGTGGCTCAGTCCCTCATTTCTATTGTAGATCG 4318  
Db 4261 GAGAAAGTCCAGTTTTTTCAGACAGTGGTGGCTCAGTCCCTCATTTCTATTGTAGATCG 4320

Qy 4319 GCCAATGAAATGCAATTAATAAACCTCTCTCAAAACAATGACAAACCATCACTGGCTTGGAGCCA 4378  
Db 4321 GCCAATGAAATGCAATTAATAAACCTCTCTCAAAACAATGACAAACCATCACTGGCTTGGAGCCA 4380  
Qy 4379 TACACCAAGTATGAGTGTAGAGTCTTAGCTGTGGAATATATGGCTGGAAAGTGTCTTCTGCC 4438  
Db 4381 TACACCAAGTATGAGTGTAGAGTCTTAGCTGTGGAATATATGGCTGGAAAGTGTCTTCTGCC 4440  
Qy 4439 TGGGTCTCTCAAGAAAGACGGAGAAATCAGACCTGTATTATCATATCCCTCTCAGTCTTT 4498  
Db 4441 TGGGTCTCTCAAGAAAGACGGAGAAATCAGACCTGTATTATCATATCCCTCTCAGTCTTT 4500  
Qy 4499 CCCCTCTCTTCTGTACTCTCTCAATATCTCTGGGAGAGCCAGCAGATATATTTACAAGA 4558  
Db 4501 CCCCTCTCTTCTGTACTCTCTCAATATCTCTGGGAGAGCCAGCAGATATATTTACAAGA 4560  
Qy 4559 GGAAGATTTGTGGGTATGACATCAATATGCTTTCTGAAACAATCAGCTCAACAGTCTATT 4618  
Db 4561 GGAAGATTTGTGGGTATGACATCAATATGCTTTCTGAAACAATCAGCTCAACAGTCTATT 4620  
Qy 4619 CCCATGCGTTTTTCACAGCTGTGGACACTGTCTAAATCCCAAGAACTATCTTACACTGTA 4678  
Db 4621 CCCATGCGTTTTTCACAGCTGTGGACACTGTCTAAATCCCAAGAACTATCTTACACTGTA 4680  
Qy 4679 GAAGACTGGAACCTTTATAGGATATATAGTATTACTATTACTCTCTGCAATTCAGTTGGT 4738  
Db 4681 GAAGACTGGAACCTTTATAGGATATATAGTATTACTATTACTCTCTGCAATTCAGTTGGT 4740  
Qy 4739 TGTGTGACAGTGTCTTGGGAGCAGCAAACTTTTAGCAGCAGCAGCAGCAACCTGAGG 4798  
Db 4741 TGTGTGACAGTGTCTTGGGAGCAGCAAACTTTTAGCAGCAGCAGCAGCAACCTGAGG 4800  
Qy 4799 CCACCTCTGGTTAAAGGAATCAACAGCACAACAATCCATCTTAAGTGGTTCCACCTGAA 4858  
Db 4801 CCACCTCTGGTTAAAGGAATCAACAGCACAACAATCCATCTTAAGTGGTTCCACCTGAA 4860  
Qy 4859 GAACCTGAATGGACCTCTCTCTATATATAGCTGGAAAGAGAGAGTCACTCTTACAGCT 4918  
Db 4861 GAACCTGAATGGACCTCTCTCTATATATAGCTGGAAAGAGAGAGTCACTCTTACAGCT 4920  
Qy 4919 CTGATGACCAAGATGATGAAAGAAATCCGTTTCATAGGAAATGGGTATTGTAATTTCC 4978  
Db 4921 CTGATGACCAAGATGATGAAAGAAATCCGTTTCATAGGAAATGGGTATTGTAATTTCC 4980  
Qy 4979 AGCTCCACTCACCCAGTCAATACAGACTTCACTGGTAAAGTGGTGGTGGTGGTGGT 5038  
Db 4981 AGCTCCACTCACCCAGTCAATACAGACTTCACTGGTAAAGTGGTGGTGGTGGTGGT 5040  
Qy 5039 TAGGAGACAGAGCTTCCAAATATTTTCTATATTTTATATCCCTTTTACAAATGAATTTT 5098  
Db 5041 TAGGAGACAGAGCTTCCAAATATTTTCTATATTTTATATCCCTTTTACAAATGAATTTT 5100  
Qy 5099 TATTATACCTACTTAGAGAAATATCTAATTCAGCCCTTTGATAGTCTTTGCTGATTTT 5158  
Db 5101 TATTATACCTACTTAGAGAAATATCTAATTCAGCCCTTTGATAGTCTTTGCTGATTTT 5160  
Qy 5159 CAGCATGCTCATCTTTTGAATTTCTGGGAAAGAAAGTCAAGTGGTGGTGGTGGTGGTGGT 5218  
Db 5161 CAGCATGCTCATCTTTTGAATTTCTGGGAAAGAAAGTCAAGTGGTGGTGGTGGTGGTGGT 5220  
Qy 5219 AATAAAGATGAAGATGAAGAGCAGCCCTTATTTGGATCAAGTATGTCTTTGATTTGT 5278  
Db 5221 AATAAAGATGAAGATGAAGAGCAGCCCTTATTTGGATCAAGTATGTCTTTGATTTGT 5280  
Qy 5279 CTTTTTGTGAAGTATGTGCCAGACATGTTTCTGAAATATTATTATCACTGTCTCTGA 5338  
Db 5281 CTTTTTGTGAAGTATGTGCCAGACATGTTTCTGAAATATTATTATCACTGTCTCTGA 5340  
Qy 5339 GCAAATGAGTTTGCAAAATGCCCTCATGCTATTTGGAGATTCTCAGTATGACCCCGTTAC 5398  
Db 5341 GCAAATGAGTTTGCAAAATGCCCTCATGCTATTTGGAGATTCTCAGTATGACCCCGTTAC 5400

1019 ATTCACTTAGTGTGAGGTGATCAGACAAAATCAGCTTCTTTATCAATGCGGTGGAG 1078  
1021 ATTCACTTAGTGTGAGGTGATCAGACAAAATCAGCTTCTTTATCAATGCGGTGGAG 1080  
1079 AAGGATCATACACCTTTCAATGCAAGAACTCTAAGTGGTTCAATACAGATTTGCGATCT 1138  
1081 AAGGATCATACACCTTTCAATGCAAGAACTCTAAGTGGTTCAATACAGATTTGCGATCT 1140  
1139 GGTACTGTGCAATAGGACAGAGTTTAAATGGTTTAGAGCAGTTGTGCGAAGAAATGCAA 1198  
1141 GGTACTGTGCAATAGGACAGAGTTTAAATGGTTTAGAGCAGTTGTGCGAAGAAATGCAA 1200  
1199 GATTTTCGATTAATACCAAGTGGACCTTAACAACAGAGAAATCTGGAAGTCTTCTGGA 1258  
1201 GATTTTCGATTAATACCAAGTGGACCTTAACAACAGAGAAATCTGGAAGTCTTCTGGA 1260  
1259 GATCTTCTCAGATGTCATGCCCAATCACAATGCCGTGCGCTGCGAGCCACCGCGGGTC 1318  
1261 GATCTTCTCAGATGTCATGCCCAATCACAATGCCGTGCGCTGCGAGCCACCGCGGGTC 1320  
1319 CACCCCTTTGGCACAGCGGTACTGCAATTCCTAATGATGACGAGACACAGCTGATAATAGA 1378  
1321 CACCCCTTTGGCACAGCGGTACTGCAATTCCTAATGATGACGAGACACAGCTGATAATAGA 1380  
1379 GTGTACCGTTGAATCTGAGGCCCATCTCTCTCTTTTGTCAATGATAATGATGGT 1438  
1381 GTGTACCGTTGAATCTGAGGCCCATCTCTCTCTTTTGTCAATGATAATGATGGT 1440  
1439 ACTTCATGGTTTCAAAATGTTTACAACATTAACAAGCTTAATCAAGGAGTGAATTT 1498  
1441 ACTTCATGGTTTCAAAATGTTTACAACATTAACAAGCTTAATCAAGGAGTGAATTT 1500  
1499 TCAGTTGATTTGGAAAATGGACAGATGATCAGGTGTTTATATATCAATCAGTTCTTTAGT 1558  
1501 TCAGTTGATTTGGAAAATGGACAGATGATCAGGTGTTTATATATCAATCAGTTCTTTAGT 1560  
1559 CACAACCAACCGAAATAGGATTCAGAGAAAGAAAGAAATAGTTTAGATGGAGGAC 1618  
1561 CACAACCAACCGAAATAGGATTCAGAGAAAGAAAGAAATAGTTTAGATGGAGGAC 1620  
1619 TGGCAATATTTGCCAGGAATGTGGTCTTTTGAATGAAACAAATGGAGATTTGGA 1678  
1621 TGGCAATATTTGCCAGGAATGTGGTCTTTTGAATGAAACAAATGGAGATTTGGA 1680  
1679 AAACCTGATTTCTGCAACTGCTTCTCAGCTTTCCTCAATTTTACTTCCATATTTCCGTGGCAAT 1738  
1681 AAACCTGATTTCTGCAACTGCTTCTCAGCTTTCCTCAATTTTACTTCCATATTTCCGTGGCAAT 1740  
1739 GTCAATTTAGCATCTGACACCTGGACCAAAATTAATCGTCTGGATACAATACTTTCTAT 1798  
1741 GTCAATTTAGCATCTGACACCTGGACCAAAATTAATCGTCTGGATACAATACTTTCTAT 1800  
1799 AATACCCCATCTCTTCAAGATGTCCTTAAAGCCAGCAAAATAGGTTTCAATTTCAATGGG 1858  
1801 AATACCCCATCTCTTCAAGATGTCCTTAAAGCCAGCAAAATAGGTTTCAATTTCAATGGG 1860  
1859 CAGTACTATACAACTGAGACTGCTGTTAACTCTCAGACACAGATATTAATGCAAGTGAACGAA 1918  
1861 CAGTACTATACAACTGAGACTGCTGTTAACTCTCAGACACAGATATTAATGCAAGTGAACGAA 1920  
1919 ATCAACATTTAGTGGAGATGTCAGTGCCATGTCATGCGATTAATCTGGACACAACAAGC 1978  
1921 ATCAACATTTAGTGGAGATGTCAGTGCCATGTCATGCGATTAATCTGGACACAACAAGC 1980  
1979 CAGCCATATAGATGCTCTGCTCCAGAGAGCTTCACTGAAGGACTTCAATTTGATGCGC 2038  
1981 CAGCCATATAGATGCTCTGCTCCAGAGAGCTTCACTGAAGGACTTCAATTTGATGCGC 2040  
2039 TGCTTGCCTCTTTATAATGACAAGCTTTTCCGCCAAGGTGATCAAGTTTACGCTTTCAAT 2098  
2041 TGCTTGCCTCTTTATAATGACAAGCTTTTCCGCCAAGGTGATCAAGTTTACGCTTTCAAT 2100  
2099 TGTAAACCTTTGTCAATGCAACAGCGCAATTCCTCAAGGCTGCCATTAACAATCTCTGTAGAC 2158

2101 TGTAAACCTTTGTCAATGCAACAGCGCAATTCCTCAAGGCTGCCATTAACAATCTCTGTAGAC 2160  
2159 CCATTTCCCTTTTGTAGACATTTAGAGGGGAGAGAGGTTTGTGATGATTTGTGAGCAATAAC 2218  
2161 CCATTTCCCTTTTGTAGACATTTAGAGGGGAGAGAGGTTTGTGATGATTTGTGAGCAATAAC 2220  
2219 ACTACAGGAAGAACTGTGAGCTGTGCAAGGATTAATTTTCCGACAAAGTTGCTGCGAGAT 2278  
2221 ACTACAGGAAGAACTGTGAGCTGTGCAAGGATTAATTTTCCGACAAAGTTGCTGCGAGAT 2280  
2279 CTTTCGGCCATAGATGTTTGTGCAAACTCTGTGATCTGTGATACAGTTTGGCACTAGAAAATGCT 2338  
2281 CTTTCGGCCATAGATGTTTGTGCAAACTCTGTGATCTGTGATACAGTTTGGCACTAGAAAATGCT 2340  
2339 AGCATTTCTTTGTGATCAGATTTGGAGGACAGTGTAAATTTGTAGAGACACGTTGTCTGCGAGG 2398  
2341 AGCATTTCTTTGTGATCAGATTTGGAGGACAGTGTAAATTTGTAGAGACACGTTGTCTGCGAGG 2400  
2399 CAGTGCAATCAGTGCGCAGAAATGGAATCTTACAATCTTACAAGAGTTGGAATCTGATGCGCTGC 2458  
2401 CAGTGCAATCAGTGCGCAGAAATGGAATCTTACAATCTTACAAGAGTTGGAATCTGATGCGCTGC 2460  
2459 AGTCCCTGTAACTGCAATACCTCTGGGACAGTGGATGGAGATATTAACCTGTCACCAAAAT 2518  
2461 AGTCCCTGTAACTGCAATACCTCTGGGACAGTGGATGGAGATATTAACCTGTCACCAAAAT 2520  
2519 TCAGGCCAGTGCAGTGCAGAAAGCAACCTTAATTTGGGCTTTAGGTGTGATCATTTGCAATTTT 2578  
2521 TCAGGCCAGTGCAGTGCAGAAAGCAACCTTAATTTGGGCTTTAGGTGTGATCATTTGCAATTTT 2580  
2579 GGAATTAATTTTCCGAAAGCTTTTAATGATGTTTGAATGTTGAGCTTGAAGTGTGATTAACCTC 2638  
2581 GGAATTAATTTTCCGAAAGCTTTTAATGATGTTTGAATGTTGAGCTTGAAGTGTGATTAACCTC 2640  
2639 CATGCTCAGTGAACAAATTTCTGCAATCTCTGAGGAGTGTGAGTGCAGAAAGAAAGAA 2698  
2641 CATGCTCAGTGAACAAATTTCTGCAATCTCTGAGGAGTGTGAGTGCAGAAAGAAAGAA 2700  
2699 GCCAAGGACTTCAGTGTGACACCTGCAAGGAGACAGTGCATAAATGTTTGGAG 2758  
2701 GCCAAGGACTTCAGTGTGACACCTGCAAGGAGACAGTGCATAAATGTTTGGAG 2760  
2759 TGTAAAGGCTCTGATGTCACACAGCTGGATTCCTTCCCTGCGACTGTCTGTAAATGCTTAAG 2818  
2761 TGTAAAGGCTCTGATGTCACACAGCTGGATTCCTTCCCTGCGACTGTCTGTAAATGCTTAAG 2820  
2819 ACAGGGCAGTGCATCTGCAAGCCCAATGTTCAAGGAGACAGTGCATAAATGTTTGGAG 2878  
2821 ACAGGGCAGTGCATCTGCAAGCCCAATGTTCAAGGAGACAGTGCATAAATGTTTGGAG 2880  
2879 GGAATCTTCTACCTACGGCAAAATTAATTTCTTCTGTCTGTGCTGCTGCAACTGTGATAAG 2938  
2881 GGAATCTTCTACCTACGGCAAAATTAATTTCTTCTGTCTGTGCTGCTGCAACTGTGATAAG 2940  
2939 ACTGGGCAATAAATGCTCTCTGCTGTGAACAAATCAACAGGACAGATGCTTGGCAAA 2998  
2941 ACTGGGCAATAAATGCTCTCTGCTGTGAACAAATCAACAGGACAGATGCTTGGCAAA 3000  
2999 TTAGGGGTAAACAGGTCTTCTGCTGTAAATCAGTGTGAGCTTCAAGGTACAATTTGACCAAT 3058  
3001 TTAGGGGTAAACAGGTCTTCTGCTGTAAATCAGTGTGAGCTTCAAGGTACAATTTGACCAAT 3060  
3059 GACAATTTTCAACTGCGCAGATGTTGATGTTTCTTGGGAGACATTAACCTGGGACC 3118  
3061 GACAATTTTCAACTGCGCAGATGTTGATGTTTCTTGGGAGACATTAACCTGGGACC 3120  
3119 ATTTGTGACCCCAATCAGTGGCCAGTGTCTGTGTTGCTTAATTCCTCAAGGAGAGGTGT 3178  
3121 ATTTGTGACCCCAATCAGTGGCCAGTGTCTGTGTTGCTTAATTCCTCAAGGAGAGGTGT 3180  
3179 AATCAGTGTCAACAGGTGTTTATTAATTTCTCAGGCAATGCACTGGCTGCTGCTGCAATGC 3238

Qy	5399	TGAAATCTCCAAAAAGCATTTGTGAAGAAAGACTATTCAACATTTTGCTTAGCTAATAATCATGCCTAA	5450
Db	5401	TGAAATCTCCAAAAAGCATTTGTGAAGAAAGACTATTCAACATTTTGCTTAGCTAATAATCATGCCTAA	5460
Qy	5459	CAGATATTTGGATGTAATGTTTTCTTCTTTCTTCTCTCTGCTGTTTCCTTCTCTCTTTTTTTC	5518
Db	5461	CAGATATTTGGATGTAATGTTTTCTTCTTTCTTCTCTCTGCTGTTTCCTTCTCTCTTTTTTTC	5520
Qy	5519	ACTGTGACAACCTTAATATCTCATGTTCTATGAAGAACAATTGTGGGGAAAACTAATCCCAG	5578
Db	5521	ACTGTGACAACCTTAATATCTCATGTTCTATGAAGAACAATTGTGGGGAAAACTAATCCCAG	5580
Qy	5579	GGAAGAAGATAACTTCTCTTAAGCCAGGACTATGTTAAAGCAAGTGAGGCTCTTGTTTCGGT	5638
Db	5581	GGAAGAAGATAACTTCTCTTAAGCCAGGACTATGTTAAAGCAAGTGAGGCTCTTGTTTCGGT	5640
Qy	5639	CACAAAATTTAAAGGCCACTAAAAACCTCAGTGTTAATGTAAATTTTAATGCAATATTTTT	5698
Db	5641	CACAAAATTTAAAGGCCACTAAAAACCTCAGTGTTAATGTAAATTTTAATGCAATATTTTT	5700
Qy	5699	AAAAATGAAATCAAATGTGAAGCACATATAAAAAATATTATCAAAAGCTTAATAATAAGACA	5758
Db	5701	AAAAATGAAATCAAATGTGAAGCACATATAAAAAATATTATCAAAAGCTTAATAATAAGACA	5760
Qy	5759	GATTGAACTCTGTACAGACACAAATCTCGCCCTACCTGACCCTTACCCCTCTCGCCCTTAC	5818
Db	5761	GATTGAACTCTGTACAGACACAAATCTCGCCCTACCTGACCCTTACCCCTCTCGCCCTTAC	5820
Qy	5819	TAGTACCGCAATATTTTGGAAAGTCCCAATGACCTCTGTGACTTACAGCTTCTAATACGATG	5878
Db	5821	TAGTACCGCAATATTTTGGAAAGTCCCAATGACCTCTGTGACTTACAGCTTCTAATACGATG	5880
Qy	5879	ATTTCAATATPAGCTGTAAAAAARCTCTACTATTATGGTACACCAATTTTTCCAAATTTTTAAAA	5938
Db	5881	ATTTCAATATPAGCTGTAAAAAARCTCTACTATTATGGTACACCAATTTTTCCAAATTTTTAAAA	5940
Qy	5939	AAATTTTACAAGTATAAGATATATATTATGTAAACATCAATAAGATGTCAATTTAAATC	5998
Db	5941	AAATTTTACAAGTATAAGATATATATTATGTAAACATCAATAAGATGTCAATTTAAATC	6000
Qy	5999	ATCCATGAGAAAGTCAATTTTGGAGCAAAATAGCTAGTCTTTTAAATATTCGNATGTGAAG	6058
Db	6001	ATCCATGAGAAAGTCAATTTTGGAGCAAAATAGCTAGTCTTTTAAATATTCGNATGTGAAG	6060
Qy	6059	ACAATGAAATGGAATTCGAGCTATAAAAAATTTGTATTTGTTTTTATTTTAAATAATAGT	6118
Db	6061	ACAATGAAATGGAATTCGAGCTATAAAAAATTTGTATTTGTTTTTATTTTAAATAATAGT	6120
Qy	6119	AAATAGTTTGCTTTTCAATTTGAGACTGGCTGCTGATGCACTTTGGTAAATGAATCATGATTA	6178
Db	6121	AAATAGTTTGCTTTTCAATTTGAGACTGGCTGCTGATGCACTTTGGTAAATGAATCATGATTA	6180
Qy	6179	TATTTCTAACCTGAGATATATTGAGATTAATGCAATGATTAACCTACTCTCTCAGTACATCAA	6238
Db	6181	TATTTCTAACCTGAGATATATTGAGATTAATGCAATGATTAACCTACTCTCTCAGTACATCAA	6240
Qy	6239	ATCATTGCAAGATTTAGAAAATTTGAACCAATGAGCTTAAAAATAGCTCAACTCTCTGCTTTAT	6298
Db	6241	ATCATTGCAAGATTTAGAAAATTTGAACCAATGAGCTTAAAAATAGCTCAACTCTCTGCTTTAT	6300
Qy	6299	ATTTCTTAAATGCGAAAAAATTTTTAAAAA	6330
Db	6301	ATTTCTTAAATGCGAAAAAATTTTTAAAAA	6332

SECRET

REF ID: A6151717

## LOCUS

## DEFINITION

**ACCEPTED**

ACCESSION  
VERSION

## KEYWORDS

**SOURCE**  
**ORGANISM** Mus musculus (house mouse)  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus; 1 (bases 1 to 5126)  
**AUTHORS** Huang, D., Eudy, J.D., Uzelvolygi, E., Davis, J.R., Talmadge, C.B., Pretto, D., Weston, M.D., Lehman, J.B., Zhou, M., Seemayer, T.A., Ahmad, I., Kimberling, W.J. and Sumegi, J.  
**TITLE** Identification of the Mouse and Rat Orthologs of the Gene Mutated in Usher Syndrome Type IIA and the Cellular Source of USH2A mRNA in Retina, a Target Tissue of the Disease  
**JOURNAL** Genomics 80 (2), 195-203 (2002)  
**MEDLINE** 2150873  
**PubMed** 12160733  
**REFERENCE** 2 (bases 1 to 5126)  
**AUTHORS** Eudy, J.D., Talmadge, C.B., Weston, M.D., Yao, S.-F., Cosgrove, D., Ahmad, I., Kimberling, W. and Sumegi, J.  
**TITLE** Isolation and Characterization of the Murine Homolog of the Usher Syndrome Type 2A Gene  
**JOURNAL** Unpublished  
**REFERENCE** 3 (bases 1 to 5126)  
**AUTHORS** Eudy, J.D., Yao, S.-F., Weston, M.D., Kimberling, W. and Sumegi, J.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (14-MAY-1999) Pathology / Microbiology, University of Nebraska Medical Center, 985454 Nebraska Medical Center, Omaha, NE 68198-5454, USA  
**FEATURES**  
**source**  
1..5126  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/chromosome="1"  
226..4611  
/notes="Usher syndrome type 2A ortholog"  
/codon\_start=1  
/product="putative extracellular matrix protein MUSH2A"  
/protein\_id="AA070550.1"  
/db\_xref="GI:7920157"  
/translation="MHYLAISPGFLCYTKTLLILAVLSVLAAQGVPRLENVCA  
FRKVTPTHTATCGPGSTFCGSPVAAEHVQLRTERLCIQDCPVRSASPLYTALLEG  
LRSCIPADDDGLHPYSSSVSPMGSHQNCPSLRAPRLAAELTIAVLWLEGGTMC  
VIEKTVGQIVFVKTISDKEITFYRTVGLQPIIKMTPGRIILMRKWLHTLVQVQHT  
ALSFVVLGSESTAFPHLTIQPHCRCPGSHPRVHPSVQYICPNAGQDFPHRMSRLNP  
LNTRELYFGSDGDFHLHIQPHCRCPGSHPRVHPSVQYICPNAGQDFPHRMSRLNP  
EAHPLFSINDDDVATSWIGFVNTIOLYGVGVAISIDLENGQXVLKVIYTPSSLOPV  
AIRIQRKADSSPWRDQYFARNCSVMGMEDNEDLNPNVNCQLQLEDFIPSHGNVT  
FOLLTSGQRHPCYNDPYNSSVLOEPMRATQIRLHPHGQYIPAGHTVDMRHQYIAYDE  
IIVSGRCQCHGAETCDYRATPPCLSPHSFTEGQDCRCSPLNDKPRSGDNVA  
FNKPCQCHGASCHYDASVDPFPLEHNRGGGVGDCQHHITGHSCSCDQYFVRS  
VADPAAPACKLCDNCRAGTNGSLHCDPIGQDCRKRVSRCQLQCDQDFYDLQA  
LDPGCRCPNCPSTGVDGIITCHONGSCQCKANVIGLRNCRNCPKFLQSPNGDG  
CSPCNLHGSVNOLCDPLSGQACCKEAKGLKDCSCRENFYGLPWSACEVDCDSKAG  
SOPGTQDTETQCTQCKENPVGRQSCQKAGFNLQNDSHLCLTNCRSGMTVNGSL  
RDXSTGQCPCKLGVGLKCHQCKRHPNLTMDNPGQCAQCDSDGLTLPGSMCDPIS  
GCCLLPHRQGRHCQOCQPGFYSSPNATGCLPCLCTAGAVSHI CNSVTGQCSKDP  
STTGSCHQCSQYRFDLPTRCPCHVAGASNGTCDVATGQCFCKBFLVGTSGKCD  
TCVPGASHLDVNNLGLKTPQCPPRGWQSSSTINVSWSPPCEPCNAHMLTYTLFR  
NGSEIYITDEHPHYTQYELDTSLAPHTAYSYIETSNVHSSSTIPRIYETKPEVSE  
CHNLTHIIPVGSDSITLTWGLSSSDPVAKYVLSCITPVDSTPCVSYEGPETSATI  
WNLVFTQYCFSGQCNESCFSYLIPIVTTAAQPPQTQGPPTVTKSLTELRIEWS  
PVDNSGIIILSYELMYRMLLTSESLVFESHGSHSALQSNVPSKLLQQAQSTFI  
SGLEPHTYAFRLAVNMGARVSSAWASERTGESVFMAPPSPVSLSPHSLSVSWEK  
PAENFTRGRIIGYKLSMVSEHPFLADVPVMCKMVHPAKSQDQSYIVRGLEPYRTYF  
TVSLCDSVGCVTSAALSGGQTAAAG"

**BASE COUNT** 1269 a 1291 c 1228 g 1338 t  
**ORIGIN**

43.4%; Score 2748; DB 10; Length 5126;  
Best Local Similarity 76.0%; Pred. No. 0;  
Matches 3512; Conservative 1; Mismatches 1066; Indels 45; Gaps 8;

169 TCATTAAGAAGTATGATACGAGCTACTCAATGTTTCGCAATTCGTAAGACGTCGTT 227



QY	2384	CACGTGTCTGGCAGGCGAGTGC	CAATCAGTGCAGAAATGGATTTCT	CAATCTCTACAGAGTTG	2443
DB	2215	CGTGTGTCTTGGCAGACAGTGC	CTTCAGTGCAGAGTGAATCT	TATGACCTTCGACGGCGCTA	2274
QY	2444	GATCTGTAGTGGCTGCGAGTCC	CTGTAACTGCAATACCTCTCGG	GACAGTGGATGAGATAT	2503
DB	2275	GATCTGTAGGCGCTCGCGTCC	CTGTAACTGTAACTCCCTCTGG	ACCGTGGATGAGACAT	2334
QY	2504	ACTGTGCACCAAAATTCAGGCG	CAGTGCAGTGCAGAAACAA	CGTTATTTGGGCTTAGGTGT	2563
DB	2335	ACTGTGCACCAAAATTCAGGCG	CAGTGCAGTGCAGAAACAA	CGTTATTTGGGCTTAGGTGT	2394
QY	2564	GATCATTTGCAATTTTGGATTT	TAATTTCTCGGAGCTTTTAAT	GTATGATGTTGGATGTGAGCCC	2623
DB	2395	AATCGCTGCATTTGCGATTTA	ATTTAATTTCTCCAGAGTTTT	TAATGGTGTGATGATGTGAGCCT	2454
QY	2624	TGCCAGTGTAACTCCATGGCT	CAGTGCAGAAACAAATTTCTG	CAATCTCTCACTCTGGGCGAGT	2683
DB	2455	TGCCAATGTAACTCCACGGCT	CAGTGCAGAACCACTCTGTGT	ATCCGCTTTCTGGGCGAGTGT	2514
QY	2684	GAGTGCAGAAAGAGCCAAAGCA	CTTCAGTGTGACACCTGCGAG	AGAAAACTTTTATGCG	2743
DB	2515	GCATGCAGAGAGGAGCCAAAG	CACTTAAGTGTGACTCTTCTG	CAGAGAAAACCTTTATGCG	2574
QY	2744	TTAGATGTACCAATTTGTAAAG	CGCTGTGACTGTGACACAGT	GTGATCCCTCTGGGAGT	2803
DB	2575	TTGCCCTGAGTGTCTTGTGAG	GTCTGTGACTGTGACAGCAAG	CGCTGGTCCCGAGCTGGGACT	2634
QY	2804	GTCTGTAACTGCTAAGACAGG	CAGTGCATCTGCAAGCCCAAT	CTTGAAGGGAGACAGTGC	2863
DB	2635	GTCTGTGACACTGAGACAGG	CAGTGCCTTTTGAAGCCCAAT	GTGTGAGGGAGACAAATGT	2694
QY	2864	AATAAATGTTTGGAGGAGAA	CTTTACCTACGGCAAAATAAT	TTCTTTCTCTGTCTGTCCT	2923
DB	2695	AGTCAGTGTAAAGCGAGGCT	ACTTCAACCTATATCAGATG	ATTTCCACCTCTGTGTGACT	2754
QY	2924	TGCACTGTGATAAGCTGGGCA	ATAATTAATGCTCTCTGCTG	TGTAACTCAAACTCAACAGGA	2983
DB	2755	TGTAACTGCGAAGAGATGGGA	ACGGTAATGGCTCTCTAAGG	TGTGACAGTCAACCGGG	2814
QY	2984	CAATGCTCTTGCATTAAGGGT	TAAAGCTCTTGGCTGTAACT	CAAGTGTGAGCTCTACAGG	3043
DB	2815	CAGTGTCTTGCATTAAGGGT	TAAAGCTCTTGGCTGTAACT	CAAGTGTGAGCTCTACAGG	2874
QY	3044	TACAAATTTGACCAATTTGAC	CAATTTTCAACTGCGAGATG	TGTGAGTGTGATCTCTGGGG	3103
DB	2875	TTCAATTTGACAGTGAACAAT	CTCTAAGTGTGACGGTGGAC	TGTGACTGTGCTGTGGGG	2934
QY	3104	ACATTAACCTGGGACCAATTTG	ACCAATTCAGTGGCCAGTGC	CTGTGTGTGCTCTAATCGT	3163
DB	2935	ACATTAACCTGGGACCAATTTG	ACCAATTCAGTGGCCAGTGC	CTGTGTGTGCTCTAATCGA	2994
QY	3164	CAAGGAAGAGTGTAACTAGTGT	CAACAGAGTTTTTATTTATTT	CTCCAGGCAATGCCACT	3223
DB	2995	CAAGGAAGAGTGTGTGAGCAG	TGTGACCGAGGCTTTTACAG	TCTCTCCAGCAATGCCACT	3054
QY	3224	GGTCCCTGCCATGCTCATGCC	ATACACTGCGGCGAGTTAAT	CACATCTGTATATGCGCTG	3283
DB	3055	GGTCCCTGCCATGCTCATGCC	ATACACTGCGGCGAGTTAAT	CACATCTGTATATGCGCTG	3114
QY	3284	ACTGTCTAGTGTGTTGCCAAG	ATGCTTCCATTTGCTGGGCA	AGGTTGTGACCAATGCAGAA	3343
DB	3115	ACGGGACAGTGTCTTGGCAAT	GACCCCTCAACCACTGGGCG	TAGTGTCAACAGTGTCAA	3174
QY	3344	GACCAATTAATTTGGATTTG	ATCCTCAGACTGGAAGATGT	CAGCCTTTGTAAATGTCACTC	3403
DB	3175	GAAAGTAACTTTAGATTTG	ATCCCTGACTGGAAGATGC	AGGCGCTTGTCACTGTCAATGA	3234
QY	3404	TCAGGAGCCTTGAATGAACCT	GTCACTTGGTTCAGAGCCAG	TGTCTGTAAACAATTT	3463
DB	3235	GCAGGAGCCTTGAATGGAACT	TTGTGTATGCAAGTCAAGG	CCAGTGTCTTCTGCAAGAGTTT	3294
QY	3464	GTCACTGGCTCAAAAGTGTG	ATGCTTGTGTTCCAGTGC	AGGCCACTTGTGATCTCAACAAT	3523

DB 3295 GTCACTGGCTCAAAATGTGCACATTGTTTCTGGTGCAAGCCACTTGGAGCGTCAACAAT |||||

QY 3524 CTATTGGGTTGAGCAGAAAACCTCATTCACAGCACCTCCGCCAGAGAGACAAAGTTCAAAGT 3583

DB 3355 CTAATTAGGCTGAGCAAAAACCTCATCTCAGCAACCTCCACCAGAGAGATGGGTTCAAGT 3414

QY 3584 TCTTCTGCTATCAATCTCTCTGGAGTCACACCTGAATCTCCAATGCCCACTGGCTTACT 3643

DB 3415 TCTTCTACCATCAATGTGTCTGGAGTCACCCGAGTGCCCAATGCTCACTGGCTTACA 3474

QY 3644 TAACAGTTTACTCAGGGATGGTTTTGAAATCTACACAAAGAGGATCAATACCCATACAGT 3703

DB 3475 TACACTCTGTTTCAAGAAATGGCTGAGATCTACACAACCTGAAGATGAACACCCGTACTAT 3534

QY 3704 ATTCAATACTCTTAGACACAGACCTGTTAACATATACCAATATTCCTTATTAATTTGAG 3763

DB 3535 ACTCAGTACTCTTAGACACTAGTCTGTCCACACACAGACATATTCATATTAATATAGAG 3594

QY 3764 ACCACCAATGTCAATGTTTCAACAAGGAGTGTAGCTGTCACTTACAGAGACAAAACGAGG 3823

DB 3595 ACCTCAATGTGCAGTTTGCAGAGGAGCTACCTGTATTATGAGACAAAGCCAGAG 3654

QY 3824 GTCCAGAGGGAACTTGACTTTAAGTTATATCAITTCCTATTGGCTCAGACTCTGTGCACA 3883

DB 3655 GTCTCAGAAGGACACTTGAACCTTGACTCACATCATCTCTGTGGGCTCGGACTCATTTACA 3714

QY 3884 CTTACTCTGCAACACTCTCAAAATCAATCTGTGCCATAGAGAAATATATTTTGTCTCTGT 3943

DB 3715 CTTACTCTGCACTGAGCTCTCTCAACAGCTCTGATCTGTGGGCAAAATATGTTTTATCATGC 3774

QY 3944 GCCCTTTGGCTGTGGTGCAGCCANGTGTTCCTACGAAGGTCATGAACCTCAGCTACCC 4003

DB 3775 AACCTGTGGACAGCACTGAGCCCTGTGTCTCTATGAAGGTCAGAAACCTCAGCTACCC 3834

QY 4004 ATCTGGAATCTGGTTCCATTTGGCCAAGTACGAATTTTCTGTACAGGGGTACTAGCGGG 4063

DB 3835 ATCTGGAATCTGGTTCCATTCACCCAGTACTGTGTTTTCTGTCCAGGGCTGTACTAATGAG 3894

QY 4064 GGCTGTTTTACAGCTTGCCCATTAACAGTGACACAGCCAGCCGCTCCCAAAAGACTA 4123

DB 3895 AGCTGTTTCTACAGCTTGCCCTATTATAGTAACACAGCCCAAGCACTCTCTMAAGCAG 3954

QY 4124 AGTCACCTAAGATGCAAAAAATCAGTTCTACAGAACTTCATGTAGAATGGTCTCAACCA 4183

DB 3955 GGGCCACCTACAGTATGAAAAATCAGTCCCACAGAACTCAGAAATAGATGGTCTCTCTCCA 4014

QY 4184 CGCGAACTTAATGGAATAATTTAAGATATGAACTATACATCAGAGAGACTGAGATCTACT 4243

DB 4015 GTGGACTCAAAATGGTATAAATTTAAGCTATGAGTTTATATGAGGAGATGGGTCTCCACC 4074

QY 4244 AAGAAACCAATCTGAGGAAAGTCAGTGTTCAGAGCAGTGTGTGGCTCAGTCTCTCAT 4303

DB 4075 GAAGAAAGC-----CTAGTGTGTGAGCCACGGTTTGTCTAGTTCTCTAC 4119

QY 4304 TCATTTGTAGAAATCGGCAATG-----AAAATGCAITAAAACTCTCCAAAACAATGACA 4357

DB 4120 TCAGCTTTACAGTCCGTCAATCCGAGTAAGAACTCTCCAGACGCCCAAGCCAGCAG 4179

QY 4358 ACCATCACTGGCTTGGAGCCATACACCAAGTATAGTTTCAGAGTCTTAGCTGTGAATATG 4417

DB 4180 TTCAATCTCTGGCTTGGAGCCACACACAGAGTATGCAATTTAGGCTCTTGGCTGTGAATATG 4239

QY 4418 GCTGGAAATGTGTTCTCTGCTGGGTCTCAGAAAGACGGGAGATCAGCACTGTATTTC 4477

DB 4240 GCTGGCAGGGTGTGCTCTGCTGGGCTCTCAGAAAGACGGGGGAATCAGTACCCGTATTTC 4299

QY 4478 ATGATCCCTCTTCAAGTCTTTCCCTCTCTTGTACTCTCTCAATATCTCTCTGGGAGAG 4537

DB 4300 ATGGCCCTCTCTTCAAGTCTCTCACTCTCACCACACTCCCTCAGTGTCTCTCTGGAGAG 4359

QY 4538 CCAGCAGATAATGTTTACAGAGGAAAGTTGTGGGGTATGACATCAATATGCTTTCTGAA 4597

Db 4360 CCAGCAGAAAACCTTTTCAAGAGGAGAAAATTATAGGCTACAGATCAGCATGCTTCTGAA 4419  
 QY 4598 CAATACCTCAACAGCTTATCCATGGGCTTTTTCACAGCTGTTCACACTCTCTAAATCC 4657  
 Db 4420 CATTTCCCTCTACAGATGTTTCAGTCATGTCTCAAGATGGTCAATTTTCTTAAGTCC 4479  
 QY 4658 CAAGAACTATCTTACATCTAGAGGACTGAAACCTTATAGGATATATAGGTTTACTATT 4717  
 Db 4480 CAAGACCACTTATACATCTAAGGGGACTCGAGCCTTATAGGACCTACAGCTTACAGTG 4539  
 QY 4718 ACTCTCTCAATTCAGTTGGTTGTGTGACAGCTGCTTCGGAGCGAGCAAACTTTAGCA 4777  
 Db 4540 TCCCTCTCGATTCGGTGGCTGTGTAAACAGTCTTGGGATCATGAGCAAACTTTAGCA 4599  
 QY 4778 GCAG 4781  
 Db 4600 GCAG 4603

## RESULT 3

AY077844  
 LOCUS 6146 bp mRNA linear ROD 12-AUG-2002  
 DEFINITION Rattus norvegicus usherin (Ush2a) mRNA, complete cds.  
 ACCESSION AY077844  
 VERSION AY077844.1 GI:22212215  
 KEYWORDS  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

## REFERENCE

1 (bases 1 to 6146)  
 Huang, D., Eudy, J.D., Uzvolgyi, E., Davis, J.R., Talmadge, C.B.,  
 Pretto, D., Weston, M.D., Lehman, J.E., Zhou, M., Seemayer, T.A.,  
 Ahmad, I., Kimberling, W.J. and Sumegi, J.  
 Identification of the Mouse and Rat Orthologs of the Gene Mutated  
 in Usher Syndrome Type IIA and the Cellular Source of USH2A mRNA in  
 Retina, a Target Tissue of the Disease  
 Genomics 80 (2), 195-203 (2002)

## JOURNAL

MEDLINE 22150873  
 PUBMED 12160733

## REFERENCE

2 (bases 1 to 6146)  
 Sumegi, J., Huang, D. and Davis, J.R.  
 Direct Submission  
 Submitted (31-JAN-2002) Center for Human Molecular Genetics,  
 University of Nebraska Medical Center, 985454 Nebraska Medical  
 Center, Omaha, NE 68198-5454, USA

## FEATURES

source

1. .6146  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 /chromosome="13"  
 /map="between d13rat49 and d13rat76"  
 1. .6146  
 /gene="Ush2a"  
 183..4721  
 /gene="Ush2a"  
 /note="USH2A; extracellular matrix protein; similar Homo  
 sapiens Usher syndrome type IIA protein"  
 /codon\_start=1  
 /product="usherin"  
 /protein\_id="AA178289.1"  
 /db\_xref="GI:22212216"

## gene

CDS

/translation="MYLALSSGFLGQAIKTSILAYLAVLLAASQGFPRLENVGF  
 KSVTPGHATGPGSTFCFSAVAHQAQLCAERLCQDCPYRSASPPYFALIEGL  
 RSCIPADGDLHPYRSNSTSIFGSHKNCPSLOAPLAAEFTLAVLWLPKRGSTMCV  
 LEKTADGQIVFVKTISRETPMYRTVNGLOPIKVMTPGRLMKKWIHTVOVHETE  
 VSSFDVGLSESTAFDTRLDSINDSVPTVLQSLNGSELFGVRQDRLNLSL  
 TNREILFSGDLPHLHQSKRCPSGSPRHPVSQQICIPNGVEDITLQHRVRLNPE  
 APLPSFINDDDVATVSHVFTDITLQNGVAISIDLNGQVQVQPIETFRSSPPQVA  
 MRIOKKADLSLWEDQYFARNCSVMGMKNGDLENPNVNCLOPPIETFRSHGNVTF  
 DLTSGQKHPGDIYFNSSLLQEFWTATQIRLYFRGLFYPAWHTVDSRHRVYAVEI

TTIGRCOCHGHAECTDRTRPYRCLSPHSPTEGPGRCSPLYNDKPPRSNKNVHAP  
 NCKPCQCHGHASCHYDASMDPPPLEYNRGGGVCDDCHHTTGNCESSCQDYFPI  
 GADPADPEVCKHCNCRDGTNGSLLDLVGDQCDKRRVSRGRCPCHIGFYGLQAL  
 DPDCRCDNCPSTVDGDDITCHNSGQCSKANVIGLRCDRCSPFPKRLSLNADGC  
 EPCNCLRHSGVNLQCDPLSGQCVCKBAKGLRCDVCRNFYGLPWSACEVCDNCRAGT  
 OAGTVDAECQCVCKPSVGRGRSECKEYFNLRQNDLSLCLPCNCKTGTNGSLL  
 CDKSTGQCPKLGVTGLRCHQCKPRFNLWMDNPQCOACDCDCLTGLSGMCDPVSGL  
 QCLCLPHROBERRCVQCGCYSSPNATGCLPCLCHTVAIKNCLNSVGHYCPDPDS  
 TTGLSHQCDQRYFRPDLRCPCHCHVAGASNGTCDATGQCFCKPFTVSGKCDT  
 CVPASHUDVNLACSTPQQPPRGRVSSSAINLSPFPFPNAAHMLTITLFRD  
 DSEYITTDHPTTYTFLTSLSPHTASYIETISNVHSSTRIPSPKTPESGE  
 HLNIETHIPVADSIITLWTLGSLNHSPIEKYVLSCTPVDHTEPCVSEGETSATIR  
 NLVPTQVCFSVQCGTNGSLSSPITVTAQAPQROBPPTWKISPTLKVESWRP  
 VDSNGVILRYELVYMKRWPSTEEISLVFESHGWFHPSPASPSANQENVLQDQVSTVLS  
 GLDPHTEYAFRLVAVNAGSVSSMASERTGESAPVMAAPSVPLSPYSLSVSEKPK  
 ABNFTREGIIGIKLSMVBERSQKRDVPMCSKLVHFAESQDSYIVQRLKPYRTYSFT  
 VSLCASVGCVTSLAGEGGTLTAGKNVLTNTKCKTHVGNQYRSVAVLWASVMSLFTAP  
 SSDLDVVYLESFQKQNH"

BASE COUNT 1553 a 1520 c 1425 g 1634 t 14 others  
 ORIGIN

Query Match 42.4%; Score 2683.6; DB 10; Length 6146;  
 Best Local Similarity 75.1%; Pred. No. 0;  
 Matches 3440; Conservative 2; Mismatches 1106; Indels 34; Gaps 6;

QY 207 GCCATTGCTAAGAACGTCGTTGGTATTACCTTACTCTGAGAACGTCGTCAGTTTCCAG 266  
 Db 3 GTCTCTCTAAGAACGTCGTTGGTATTACCTTACTCTGAGAACGTCGTCAGTTTCCCA 62  
 QY 267 AAAATGGAGTATCGCAACATCATCTTAAAGTACCTCTTCAAGAGTATGCTGGCAAGTG 326  
 Db 63 GAGATGAGAGATCTCTGTATTCAAGTCCCGACGCTGGATCCAGTTCATAGTGCACAG 122  
 QY 327 CGTGGCTGATTTATTTATTTAGAAATGCTTTATCAGAGAGAGATGCTTTT-TGTAAAC 385  
 Db 123 AGATCCCTGGACAGTAGTTTAGGAAAGCTTTTCATCAGGTGAAGAATACTTTTCTGTCTC 182  
 QY 386 ATGAATGCCACAGTCTTTTCATTCGGCTCTGGCTCTTGTGTTTCAGTCAATGAATGTTG 445  
 Db 183 ATGATTACCTGGCTCTCTCA-----TCGGCTTCTGGGTGAGCCATTAACAATCG 236  
 QY 446 ATCTTTGCTATTTTGTCTTCAATATCTGTATCTGAGTCAAGAGTCTTTTCCCAAGGCTG 505  
 Db 237 ATCTTTGCTATTTTGGCTCAGTGTGTGCTGCTGCTCGCAGGGGCTTTTCCCAAGGCTG 296  
 QY 506 GAGAACGTTGGAGCTTTCAAGAAAGTTTCCATCGTGCACACCCAGAGAGTATGTGGACTC 565  
 Db 297 GAGAACGTTGGTGTCTTTCAAGAGGTATCCATTTGTGCAAGCCACGCCACGCTGGATAC 356  
 QY 566 CCAGACCGAAGCAGCTTTTGTGTCAGAGCTCTGCTGCTGCTGAAGTATTCAGTCTCTGATCC 625  
 Db 357 CCAGCCCCCAGCAGCTTTTGTGCTAGCGCGCTGGGAGCTGAGCATGCGCAGTTGTGTGC 416  
 QY 626 CAGCGGTTTGTATTTCAGAGTATGCCATACAGATCTTTCACACCTTACCTTACACTGCCCTT 685  
 Db 417 GAGAGGCTGTATTCAGAGCTGCTCCCTTACCGCTCTGCTCTCCYCCCTATACGCGCC-- 474  
 QY 686 TTCTCAGACGCGCTCAGTAGCTGATCATACACAGAGAGATGATCTGCATCTTAAACGCC 745  
 Db 475 -TCCTAGAGGCGCTCAGGAGCTGCATTCCTCAGACACCGGTGATCTCCACCCCTTACTCC 533  
 QY 746 CATAGCAATTCGCAAGTATTTTATTTTGGAAATCAGAGAGCTCTTTTCTTCTCTCTCT 805  
 Db 534 AGAGCAACTCTCAGAGCTTCATATTCGAAAGTCAAGAAGTCTGCTTCTCTCTCTCTCT 590  
 QY 806 TCTCCAAAGCTGATGGCATCAATTTTACCTTACCTTGTATGGCTTGAACCTCAGGCAACAAG 865  
 Db 591 GCTCCGAGCTGGCAGCAGATTTTACCTTAGCCGTGTGGTGTGAAGCCCGAGGAGGCGCAG 650  
 QY 866 GTAATGTGTGTTATAGAAAAGACGTGATGGGAGATGTGTTGTTCAAACTTACAATATCT 925  
 Db 651 ACTATGTGTGTGTAGAAAAGACAGCGGATGGGAGATGTGTTGTTTCAAAAGTTTACAATATCT 710



Db	2871	GACTGTGACTCGTTGGGGACANATTACTTTGGGAAGCAATGTGTGACCCCGTCAGTGCCCACTGC	2933
Qy	3146	CTGTGTGTGCTCTAATTCGTCAAGGAAGAGGTGTAATCAGTGTCTCAACACAGGTTTTTATATT	3205
Db	2931	CTATGCCCTACCTCATCGTCAAGAAGAAGGTGTGTACAGTGTCAACACAGGGTTTACAGT	2990
Qy	3206	TCTCCAGGCAATGCACTGGCTGCTGCCAFGCTCATGCCANTACAACCTGGCGCATTTAAT	3265
Db	2991	TCTCCACAGCAATGCACTGGCTGCTGCCATGCTTATGTGCCACACAGTTGTCTACCAAGAA	3050
Qy	3266	CACATCTGTAATAGCCTGACCTGGTCAGTGTGTTGGCCAAAGTCTTCATTTGCTGGGCAA	3325
Db	3051	TGCATCTGTATATAGTGTACGGGACATTTGCTATTTGGCCCTGACCCCTCAACCACTGGGCTG	3110
Qy	3326	CGTTGTGACCAATGCAAAAGACATTACTTTGGATTTTGATCTCTCAGACTGGAAGATGTCAG	3385
Db	3111	TCATGGCACCAAGTGCCAAGATAGTACTTTAGATTTTGATCCCTGACTGGAAAGATGCAGG	3170
Qy	3386	CCTTGTAAATTTGTCATCTCTCAGGAGCCTTGAATGAACCTGTCTACTTTGGTCAAGGCCAG	3445
Db	3171	CCTTGTCACTGTCTATGTAGCAGGAGCCTCGAATGGAACCTTGTGATGCAGTCAAGGCCAG	3230
Qy	3446	TGTTTCTGTAAACAATTTGTACCTGCTCAAGTGTGATGCTTGTGTTCCAGTGCAGAC	3505
Db	3231	TGTTTCTGCAAGAAGTTTGTCACTGGCTCAAAATGTGACACTTGTGTCTGGTGCAGAC	3290
Qy	3506	CACTTGATGTCAACAATCTATTGGGTTGCAGCAAAACTCATTTCCAGCAACCTCGGCC	3565
Db	3291	CACTTGACGTCAACAATCTATTAGCTTTGTTCAAAACTCACTCAGCAACCTCCACCC	3350
Qy	3566	AGAGGACAAGTTCAAAGTTCTTCTGCTPATCAATCTCTCTGGAGTCCACCTGATTTCTCCA	3625
Db	3351	AGAGGACGGGTTCAAAGCTCTTCTGCCATCAATCTCTCTGGAGTCCACCTGATTTCCC	3410
Qy	3626	AATGCCCACTGGCTACTTACAGTTTACTCAGGGATGGTTTGAATCTACACAACAGAG	3685
Db	3411	AATGCACTCTGGCTTACATACACTCTGTTCAGGGATGATTTCCGAGATCTACACAACCTGAC	3470
Qy	3686	GATCAATACCATACATACAGTATTCAATACCTCTTAGACACAGACCTGTTACCATATACCAA	3745
Db	3471	GATCAGACACCTTACTATACTCAGTACTCTTTGGACACTAGTGTGTCAACCGCACAGCA	3530
Qy	3746	TATTCTTATTAATGTAGACCAACAAATGTGCTAGTGTTCACAAGAGAGTGTAGCTGTCACT	3805
Db	3531	TATTTCAATATACATAGAGACCTCCAAATGTGCACAGTTCAACAAGGACATTCCTGTGATT	3590
Qy	3806	TACAAGCAAAAACGAGGGTCCAGAGGGAAAATTGACTTTTAAAGTTATATATCTTCCTATT	3865
Db	3591	TACAAGCAAAAGCCAGAGGGCTCAGAAGGCCACTTGAACCTGCACTCATTCATTTCTGTG	3650
Qy	3866	GGCTCAGACTCTGTGACACTTACCTGCAACAACACTCTCAAAATCAATCTGGTCCCATAGAG	3925
Db	3651	GCCTCAGACTCCATTACACTTGTCTGACCTGGACTCTCCAAACCACTCTGGTCCCTATAGAG	3710
Qy	3926	AAATATATTTTGTCTGTGCCCTTTTGGCTGGTGGTCAGGCATGTGTTTCTTACGAAGGT	3985
Db	3711	AAGTATGTTTTATCATGCACACCTGTGTGGAACCACTGAGCCCTGTGTCTCTTATGAGGT	3770
Qy	3986	CATGAACCTCAGCTACCACTGGAATCTGGTTCATTTGGCCAAAGTACGAATTTTCTGTGA	4045
Db	3771	CCAGAGACCTCAGCTACCATCAGGAATCTGGTTCATTTCCACCCAGTACTGTTTCTCTGTG	3830
Qy	4046	CAGCGGTGTAAGCGGGGCTGTTTACAGCTGTGCCATTAAGTGTGACCAAGGCCAG	4105
Db	3831	CAGGCTGTACTATGAGGAGCTGTTTATATAGCTCACCTATTACAGTAAACAGGCCAA	3890
Qy	4106	GCCCTCTCCCAAGACTAAGTCCAACCTTAAGATGCAAAAAATCAGTTCTACAGAACTTCAT	4165
Db	3891	GCGCTCTCCCAAGGACGAGGCCACTCAGTATGGAATAATCAGTCCACAGAACTCAA	3950
Qy	4166	GTAGATGTTGTTCTCACACAGCGGAACTTAAATGGAATAATTAAGATATGAATATACATG	4225
Db	3951	GTAGATGTTGTTCTCGGCCAGTGGATTCAAAATGGTGTAAATTAACAGTATGAGTTATACATG	4010

QY	4226	AGAAGACTGAGATCTACTAAAGAAACACAAATCTGAGGAAAGTGCAGTTTTTCAGAGCAGT	4285
Db	4011	AAGAGATGGCGCTCCA-----CAGAGGAGAGTCTCGTGTTCAGAGCCAT	4055
QY	4286	GGTGGCTCAGTCTCTCAATTCATTGTAGAAATCGGCCAA-----TGAATATGCATTAAAA	4339
Db	4056	GGTTGGTTCANCTCTCACCCAGCTCACCGTCAGCCCAATCAGAGTGAGAAATGTGCTTCAG	4115
QY	4340	CTCTCTCAAAACAATGACAACCATCACCTGGCTTGGAGCCATACACCAAGTATGAGTTCAGA	4399
Db	4116	GATCCCAAGTCAGCACAGTCTCTCTGGCTTGGACCCACACACAGAGTATGCATTTAGG	4175
QY	4400	GTCTTACTGTGAATATGGCTGGAGATGTCTCTTTCGCTGGGTCTCAGAAAGAACGGGA	4459
Db	4176	GTGTGGCTGTGAATATATGGCTGGCAGTGTGTCACTGTGCTGGGCTTCAGAAAGAACAGGA	4235
QY	4460	GAATCAGCACCTGTATTATCATGATCCCTCTTCAGTCTTTCCCTCTCTCTCGTACTCTCTC	4519
Db	4236	GAATCAGCACCTGTGTTCATGGCTGTCTCTTCAGTCTCTCACCTCTCACCATACTCCCTC	4295
QY	4520	AATATCTCTGGGAGAGCCAGCAGATAATATGTTACAAGAGGAAAAGTTGTGGGGTATGCAC	4579
Db	4296	AGTGTCTCTGGGAGAGCCAGCAGAAAACCTTTACAAGAGAGAGATATTAGGGTACAG	4355
QY	4580	ATCAATATGCTTTCTGAAACAATCACCTCAACAGTCTATTTCCTCAATGGCGTTTTTCACAGCTG	4639
Db	4356	ATCAGCATGGTTCTTGAAACGTTCCTCTCAACGAGATGTTCCAGTCACTGTCTCAAAAGCTG	4415
QY	4640	TTGCACACTCTAAATCCCAAGAACTATCTTACACTGTAGAAAGGATCGAAACCTTATAGG	4699
Db	4416	GTGCATTTTCTGAGTCCCAGAACCAGTCTTATCATCGTTCAGAGACTCAAGCCCTTATAGG	4475
QY	4700	ATATATGAGTTTACTATTACTCTCTGCAATTCAGTTGTGTGTGACCAAGTGTCTCGGGA	4759
Db	4476	ACCTACAGTTTTTACTGTCTCCCTCTGGCTTCGTGGGCTGTGTAAACCAAGTGTCTTTGGGA	4535
QY	4760	GCAGGACAAACTTTAGCAGCAG	4781
Db	4536	GAAGGACAAACGGCTAACAGCAG	4557

```

RESULT 4
HSUSH2A17
LOCUS
DEFINITION
    Homo sapiens Usher syndrome type Iia protein gene, exon 21 and
    complete cds.
ACCESSION
    AF091889
VERSION
    AF091889.1 GI:8515085
KEYWORDS
    17 of 17
SOURCE
    Homo sapiens (human)
ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
    1 (bases 1 to 2652)
    Weston,M.D., Rudy,J.D., Fugita,S., Yao,S.-F., Usami,S., Cremers,C.,
    Greenburg,J., Ramesar,R., Martin,A., Moiler,C., Smith,R.J.,
    Sumegi,J. and Kimberling,W.J.
    Genomic structure and identification of novel mutations in usherin,
    the gene responsible for Usher syndrome type Iia
    Am. J. Hum. Genet. 66 (4), 1199-1210 (2000)
TITLE
    JOURNAL
    MEDLINE
    PUBMED
    20206315
    10729113
REFERENCE
    2 (bases 1 to 2652)
    Weston,M.D., Rudy,J.D., Fugita,T., Sumegi,J. and Kimberling,W.J.
    Direct Submission
    AUTHOR
    TITLE
    Submitted (14-SEP-1998) Genetics, Boys Town National Research
    HOSPITAL, 555 North 30th Street, Omaha, NE 68131, USA
JOURNAL
FEATURES
    source
    Location/Qualifiers
    1..2652
    /organism="Homo sapiens"
    /mol_type="genomic DNA"

```

QY 926 GAGAAAGAGACCAATGTTTTATTATCGCACAGATAAATGTTTGCACCTCCAAATAAAGTA 985  
 DB 711 GAGAGAGAGACCAATGTTTTATTATCGCACAGATAAATGTTTGCAGCTCCAAATAAAGTA 770  
 QY 986 ATGACACTGGGAGAGATTTCTGTGAAGAAATGATTCATCTTAGTGTGAGTGCGATCAG 1045  
 DB 771 ATGACACCAAGGAGAGATTTCTGTGAAGAAATGATTCATCACTGTGAGTGCGATCAG 830  
 QY 1046 ACAAATAACAGCTCTTTATCAATGCGGTGAGAGAGATCATACACCTTTCAATCCAGA 1105  
 DB 831 ACGGAAGTCAGCTCTTTGTGATGTTTGGAGGAGACAGACTGCATTCATCAAGA 890  
 QY 1106 ACTCTAAGGTTCATTAACAATTTGATCTGGTACTGTGCAATAGGACAGAGTTTA 1165  
 DB 891 ACTCTCCGGAATCAATCATGACTCAGTCCCGCAGCAGCTGACTAATAGGACAGTTTA 950  
 QY 1166 AATGCTTTAGACAGTTTGTGGAAGATGCAAGATTTTCGATTAACAAGTGGCATTT 1225  
 DB 951 AACGGTTTCAGAGCTGTTTGTGGAAGATGCAAGATTTCCGATTTGTACAAAGTGTCACTT 1010  
 QY 1226 ACAAACAGAGAGATTTCTGGAAGTCTTCTCTGGAGATCTTCTCAGATTCGATGCCAATCA 1285  
 DB 1011 ACAAACAGAGAGATTTCTGAGCTCTTTCTGGAGATCTCCCGCAGCTTGCACATTCAGTCA 1070  
 QY 1286 CATTGCGGTGCGCTGGAGCACCACCGCGGTGCCACCTTTGGCAGCAGCGTACTGCAAT 1345  
 DB 10 1 CATTGCGGTGCGCAGGCGCAGCCTCCCGGTGTCACCGCTCGGTTTCAGCAGTATTCGCAAT 1130  
 QY 1346 CTAATGATGACGAGACACAGCTGATTAATAGAGTGTCAAGTGTCAAGTTCGAGCCCAT 1405  
 DB 1131 CCCAATGTTGGAAGACACACTCCAGCATCGAGTGTCAAGCTGCAATCCTCAAGCCCAT 1190  
 QY 1406 CTTCTCTCTTGTCAATGATGATGATGTTGATCTTCAATGCTTTCAATGCTTTTACA 1465  
 DB 1191 CTTCTCTCTTCAATGATGATGATGTTGATCTTCAATGCTTTTACA 1250  
 QY 1466 AACATTACAGCTTAAATCAAGGATGATATTTTCAGTGTGATTTGGAAATGGACAGTAT 1525  
 DB 1251 GACATTACAGCTCAATCAAGGATGCGCAATCTCTATAGACTTGGAAATGGACAGTAT 1310  
 QY 1526 CAGGTGTTTTATATCATTCAGTCTTTAGTGCACAAACCAACGGAATTAAGGATTCOA 1585  
 DB 1311 CAGGTGTTTTCAATACCAATTCGGTCTTCCAGCCACACAGCCTGTGCGCATCGGATCCAG 1370  
 QY 1586 AGGAAGAGGAAATAGTTTATAGTGGGAGCTGGCAATATTTTCCAGGAAATTTGGT 1645  
 DB 1371 AGGAAGAGGAGACAGTCTCTCTGGGAGGAGCTGGCAGTATTTTGGAGAAATTTGCACT 1430  
 QY 1646 CTTTGTGAATGAAACCAATGGAGATTTGGAACCACTGATTTCTGCAACTGCTTCCAG 1705  
 DB 1431 GTTTGGGGAATGAAACCAATGGAGATCTGGAATCCCAATTTCTGCACTGCTCCAG 1490  
 QY 1706 CTTTCCAAATTTTACTTCCAAATTTCCGTTGGCAATGTCATTTAGCATCTGACACCTGGA 1765  
 DB 1491 TTTCCCGAGTTTATCCCGTTTCCCAAGTAAATGTCACCTTTGACCTCTTAACATCTGGA 1550  
 QY 1766 CCAAAATATCTCTCGGATCAATTAATCTTATATATACCTTCTTCAAGAGTCCGTA 1825  
 DB 1551 CAGAGCATCTCTCTGGGAGCTATGACTTTTACCAACAGCTCAATTTGCTGCAAGAAATTCATG 1610  
 QY 1826 AAAGCCAGCAATAAGGTTTCAATTTTCATGGGAGTACTATACACTGAGACTGCTGTT 1885  
 DB 1611 ACAGCCACTCAGATACGCTGACTTCCGAGGCTGTTTCAATCCGCTTGGCACTGTT 1670  
 QY 1886 AACCTCAGACACAGATATATGACGTGGACGAAATCACCAATTTAGTGGAGATGTCACTGC 1945  
 DB 1671 GACTCAAGACACAGATATATGACGTGGATGAGATCAACATTTAGTGGAGATGTCAATGC 1730  
 QY 1946 CATGTCATGCGGATATCTGGACACAAACAGCCAGCCATATAGATGCTCTGCTCCAG 2005  
 DB 1731 CATGGCCATGCGGAGACTGTGACAGAAACAGAGCGGCTTACCGGCTGCTCTGCTCACCG 1790  
 QY 2006 GAGAGCTTTCAGTGAAGGACTTCAATTTGATCGCTGCTGCTCTTTTATATGACAGCCT 2065

DB 1791 CACAGCTTCACTGAAGGCCCCAGTGTGTCATGCTCCCCCTCTTTTATAATGACAGCCT 1850  
 QY 2066 TTCCGCCAAGTGATCAAGTTTACGCTTTTCAATTTGTAACCTTGTCAATGCAACAGCCAT 2125  
 DB 1851 TTCCGCCAGTGTGTAACAAAGTTTCAATGATTCATGCTTCAAGCTTGTCACTGCTGCTGCT 1910  
 QY 2126 TCCAAAGCTGCCATTACAACTCTCTGTAGACCAATTTCTTTTGTAGCACTTTCAGAGG 2185  
 DB 1911 GCCAGCAGCTGCCATATGATGCTTCCATGAGCCCAATTTCCCTGCGAGTACAAACAGAGT 1970  
 QY 2186 GGAGGAGAGTGTGTGATGATTTGAGCATTAACACTACAGGAAGAACTGTGAGCTGTGCG 2245  
 DB 1971 GGCGGAGAGTGTGTGATGATGCTGAGCAACCAACAGGAGAACTGTGAGTCAATGC 2030  
 QY 2246 AAGGATTAATTTTCCGACAAAGTTGTCAGATCCTTTGGCCATAGATGTTTGGAAACCC 2305  
 DB 2031 CAGGACTACTTTTACCGACCCCAATTTGGTGGCGATCCTGCTGACCCAGAGGTTTGCAGCAC 2090  
 QY 2306 TGTGACTGTGATACAGTTTGGCAGCTAGAAATGATGATCTTTTGTGATCAGATTTGGAGGA 2365  
 DB 2091 TGTGACTGTCAACAGGAGCGGAACCTAGAAACCGGAGCCTCTCTGTGACCTGCTGTTGGAGAC 2150  
 QY 2366 CAGTGTAAATTTGTAAGAGACAGCTGTCTGCGCAGGAGTGCATCAATCAGTGCAGAAATGATTC 2425  
 DB 2151 CAATGTGATTTGTAAGAGAGCTGTCTGCGCAGAGCGTGTCTTTCGTTGCCATATTTGGGTTT 2210  
 QY 2426 TACAATTTACAAAGTGTGATCTCTGATGCTGAGTCTCTGTAACTGCAATACCTCTGCGG 2485  
 DB 2211 TATGGCTTGCAGGCGCTGGATCTCTGATTTGCTGCGCGCTCTGTGACTGTATCTCTGCGG 2270  
 QY 2486 ACAGTGTGAGAGATTTTACTGTGCAACAAATTTGAGTAAATTTTCCGAGCTTTTAAAT 2545  
 DB 2271 ACCGTGTGAGAGATTTTACTGTGCAACAAATTTGAGTAAATTTTCCGAGCTTTTAAAT 2330  
 QY 2546 GTTATTTGGGCTTAGTGTGATTCATTTGCAATTTTGGATTTTAAATTTTCCGAGCTTTTAAAT 2605  
 DB 2331 GTTATTTGGGCTTCGATCGATCGCTGAGTTCGATTTTGAATTTTAAATTTTCCGAGCTTTTAAAT 2390  
 QY 2606 GATGTTGGATTTGAGCGCTTGCAGTGTAACTTCCATGCTCAGTGAACAAATTTCTGCAAT 2665  
 DB 2391 GCTGACGATTTGAACTTGCATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 2450  
 QY 2666 CTTCACTCTGCGCAGTGTGATGCAAAAGAGCCAAAGAGCTTTCAGTGTGACACCTGTC 2725  
 DB 2451 CCACTCTCTGCGCAGTGTGATGCAAAAGAGAGCTTTCAGTGTGACACCTGTC 2510  
 QY 2726 AGAAGAAATTTTATGCTTATAGTGTCAACCAATTTGTAAGCGCTGTGATGTCAGACACT 2785  
 DB 2511 AGAAGAAATTTTATGCTTATAGTGTCAACCAATTTGTAAGCGCTGTGATGTCAGACACT 2570  
 QY 2786 GATTCCTCTCTGCGCAGTGTGATGCAAAAGAGCCAAAGAGCTTTCAGTGTGACACCTGTC 2845  
 DB 2571 GGCACCCAGCTGCGCAGCTGTGATGCTGAGACAGGCGAGTGTGATTTGTAAGCCAGT 2630  
 QY 2846 GTTGAAGGAGACAGTGCATTAATTTTGTGAGGAGAACTTTTACCTTACCTTACCTTACCTTACCT 2905  
 DB 2631 GTTGAAGGAGACAGTGCATTAATTTTGTGAGGAGAACTTTTACCTTACCTTACCTTACCTTACCT 2690  
 QY 2906 TCTTTCTCTCTGCTGCTGCACTGTGATGAGCTGCGGAGAACTTTTACCTTACCTTACCTTACCTTACCT 2965  
 DB 2691 TCCACCTCTCTGCTTACCTTGTAACTGTGAGAGACCGGAGCAGTAAACCGCTCTCTCTACTG 2750  
 QY 2966 TGTAAACAAATCAACAGGACAAATGCTTGTGCAAAATTTAGGCGGTAAACAGCTTCTCGCTGTAAT 3025  
 DB 2751 TGTGACAGTCACTGGCAGTGTCTGCTGCAAACTAGGCGGTAACTGCTGCTGCTGCTGCTGCTGCT 2810  
 QY 3026 CAGTGTGAGCTTCAAGGTAACAATTTTGAACATTTTGAACATTTTGAACATTTTGAACATTTTGAACAT 3085  
 DB 2811 CAGTGTGAGCTTCAAGGTTCAATTTTGAACATTTTGAACATTTTGAACATTTTGAACATTTTGAACAT 2870  
 QY 3086 GAGTGTGATTTCTTGGGACATTTACTGCGGACCAATTTGTGAGCCCAATTTGAGTGGCCAGTGC 3145



QY 6040 AAATATTCATATGTGAACACAAATGAAATGGAATTCGAGCTATATAAAATTTCTATTGTTT 6099  
 DB |||||||  
 1906 AAATATTCATATGTGAACACAAATGAAATGGAATTCGAGCTATATAAAATTTCTATTGTTT 1965  
 QY 6100 TATTTTCTAAATAGTAATAGTTCTTTTCTATTCAGACTGCTGCTGATGACCT 6159  
 DB |||||||  
 1966 TATTTTCTAAATAGTAATAGTTCTTTTCTATTCAGACTGCTGCTGATGACCT 2025  
 QY 6160 TGGTAATGAATCATGATATATTTCTAACTGAGATATATTGAGATTAATGATGATTAAC 6219  
 DB |||||||  
 2036 TGGTAATGAATCATGATATATTTCTAACTGAGATATATTGAGATTAATGATGATTAAC 2085  
 QY 6220 ACTCTCTAGTACATCAAAATCATTCGAGATATTAGAAATTTGAACCATTTGAGCTTAAAA 6279  
 DB |||||||  
 2086 ACTCTCTAGTACATCAAAATCATTCGAGATATTAGAAATTTGAACCATTTGAGCTTAAAA 2145  
 QY 6280 TGCTCAACTTCGCTTTATATTTCTAAATGCAAAAAAATTTAAAAA 6330  
 DB |||||||  
 2146 TGCTCAACTTCGCTTTATATTTCTAAATGCAAAAAAATTTAAAAA 2196

RESULT 5  
 AC119429/c  
 LOCUS 115940 bp DNA linear PRI 19-FEB-2003  
 DEFINITION Homo sapiens chromosome 1 clone RP4-723P6, complete sequence.  
 AC119429 AL138934  
 VERSION HTG  
 KEYWORDS AC119429.3 GI:28416171  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 115940)  
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
 Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and  
 Haugen, E.D.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 115940)  
 Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.  
 Direct Submission  
 Submitted (26-APR-2002) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 3 (bases 1 to 115940)  
 Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.  
 Direct Submission  
 Submitted (11-FEB-2003) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 4 (bases 1 to 115940)  
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
 Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and  
 Haugen, E.D.  
 Direct Submission  
 Submitted (19-FEB-2003) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 On Feb 19, 2003 this sequence version replaced gi:28301973.  
 ----- Genome Center  
 Center: University of Washington Genome Center  
 Center Code: UMG  
 Web site: <http://www.genome.washington.edu>  
 Contact: [uwgctgs@u.washington.edu](mailto:uwgctgs@u.washington.edu)  
 Drafting Center: SC  
 ----- Project Information  
 Center project name: chr-1  
 Center clone name: RP4-723P6 (sc0827)  
 ----- Summary Statistics  
 Sequencing vector: plasmid; 75% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 115940 bases at least Q40  
 Consensus quality: 115940 bases at least Q30  
 Consensus quality: 115940 bases at least Q20

Insert size: 115940; sum-of-contigs  
 Quality coverage: 13.5x in Q20 bases; sum-of-contigs

# Overlapping Sequences:

5': RP11-22M7 (UMGC:sc0603) AC093581, 2474-bp overlap  
 3': RP5-1099E6 (UMGC:sc0859) AC138024

# Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

# Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BgIII				HindIII				EcoRI			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9577	9385	1852	1831	-----	3728	-----	3770	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
5671	5768	449	<800	-----	2184	-----	2193	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
10327	10090	512	<800	-----	8065	-----	8167	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1676	1673	2814	2853	-----	486	-----	<800	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
10319	10090	1247	1234	-----	106	-----	<800	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
494	<800	4695	4783	-----	3581	-----	3414	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8213	8182	3373	3406	-----	1068	-----	1095	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1220	1169	3011	3085	-----	1629	-----	1599	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
3450	3448	804	800	-----	313	-----	<800	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
2966	2969	7841	7860	-----	117	-----	<800	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
5992	5956	1345	1337	-----	8123	-----	8167	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
325	<800	4829	4783	-----	5653	-----	5578	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
5999	5956	83	<800	-----	811	-----	821	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1075	1070	297	<800	-----	3821	-----	3770	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
9479	9385	1092	1095	-----	3464	-----	3596	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
798	815	10	<800	-----	6982	-----	6896	-----	-----	-----	-----

QY	4960	TGSGTATGTGTAATTTCCAGCTCCACCTCACTCCAGTCAATACAGACTTCACCTGTAACTG	5019
Db	826	TGGGTATGTGAATTTCCAGCTCCACCTCACTCCAGTCAATACAGACTTCACCTGTAACTG	885
QY	5020	TGTTTGACATGCTTTTATTTAGGACACGAGCTCCAAATGTCTTCTATATTTTCATA	5079
Db	886	TGTTTGACATGCTTTTATTTAGGACACGAGCTCCAAATGTCTTCTATATTTTCATA	945
QY	5080	TCCCTTTACATGAATTTTATATATACCTACTTAGAGAAATATCTAATTCAGCCCTTTGAT	5139
Db	946	TCCCTTTACATGAATTTTATATATACCTACTTAGAGAAATATCTAATTCAGCCCTTTGAT	1005
QY	5140	AGCTTTTGCTGATTTGTTTTCAGCATGTCCATCTTTTAGAATCTCGGGGAAAAAGTCCAG	5199
Db	1006	AGCTTTTGCTGATTTGTTTTCAGCATGTCCATCTTTTAGAATCTCGGGGAAAAAGTCCAG	1065
QY	5200	GTAAGTGAAGGAAAGGAAAAATAAAGATGAAGATGAAGAGACGACCTTATTTGGATCAAA	5259
Db	1066	GTAAGTGAAGGAAAGGAAAAATAAAGATGAAGATGAAGAGACGACCTTATTTGGATCAAA	1125
QY	5260	GTAATGCTTTGTATTTTGTGTAAGTATGTGCAGACGACATGTTTCTTGAAATAT	5319
Db	1126	GTATGTGCTTTGTATTTTGTGTAAGTATGTGCAGACGACATGTTTCTTGAAATAT	1185
QY	5320	TATTCACATGTTCTCTCAGCAAAATGAGTTTGC AAAATGCCCTCATGCTATTTGGAGATTC	5379
Db	1186	TATTCACATGTTCTCTCAGCAAAATGAGTTTGC AAAATGCCCTCATGCTATTTGGAGATTC	1245
QY	5380	TCAGTATGACCCCGTTACTGAAACTCCAAAAGCATTTGAAGAAAGCTATTTCAACTTTG	5439
Db	1246	TCAGTATGACCCCGTTACTGAAACTCCAAAAGCATTTGAAGAAAGCTATTTCAACTTTG	1305
QY	5440	CTTAGCTAATCATGCTTAACAGATATTTGATGTAATGTTTCTTTTCTCTCTGCTG	5499
Db	1306	CTTAGCTAATCATGCTTAACAGATATTTGATGTAATGTTTCTTTTCTCTCTGCTG	1365
QY	5500	TTTCTCTCTCTTTTTCATGTGCACAACTTAATATCTCATGTGTCTATGAGAACATTTG	5559
Db	1366	TTTCTCTCTCTTTTTCATGTGCACAACTTAATATCTCATGTGTCTATGAGAACATTTG	1425
QY	5560	TGGGAAAACTAATCCAGGGAAAAAGATAACTCTCTAAAGCCAGACACTATCGTAAAGCAA	5619
Db	1426	TGGGAAAACTAATCCAGGGAAAAAGATAACTCTCTAAAGCCAGACACTATCGTAAAGCAA	1485
QY	5620	GTGAGGCTCTGTTTCGGTGCAAAAAATTAAAGGCACTAAAAAACTCATGCTTTAATGTAA	5679
Db	1486	GTGAGGCTCTGTTTCGGTGCAAAAAATTAAAGGCACTAAAAAACTCATGCTTTAATGTAA	1545
QY	5680	ATTTTAAATGCAATATTTTAAAAATGAAAAATCAATGTGAAGCACTATAAAAAATTTATC	5739
Db	1546	ATTTTAAATGCAATATTTTAAAAATGAAAAATCAATGTGAAGCACTATAAAAAATTTATC	1605
QY	5740	AAAAAGCTTAAATAAAGACAGATTGAACTCTGTACAGCACAATCCTGCTCACTGGCCTT	5799
Db	1606	AAAAAGCTTAAATAAAGACAGATTGAACTCTGTACAGCACAATCCTGCTCACTGGCCTT	1665
QY	5800	ACCTCTCTCTGGCCTTACTAGTACCGCAATATTTTGGAAAGTCCCATGCACTCTGTGACT	5859
Db	1666	ACCTCTCTCTGGCCTTACTAGTACCGCAATATTTTGGAAAGTCCCATGCACTCTGTGACT	1725
QY	5860	TACAGCTCTTAATAGCATGATTTCCAAATATAGCTGTAAAAAACTCTACTATTGTACACC	5919
Db	1726	TACAGCTCTTAATAGCATGATTTCCAAATATAGCTGTAAAAAACTCTACTATTGTACACC	1785
QY	5920	ATTTTCCCAATTTTAAAAAAATTTTACAAAGTATAAGTATATATTTAATGTAAACTCA	5979
Db	1786	ATTTTCCCAATTTTAAAAAAATTTTACAAAGTATAAGTATATATTTAATGTAAACTCA	1845
QY	5980	TAAAGATGTTTCATTTAATCATCCATGAGAAAGTCATTTTGGAGCAAAATAGCTAGCTTTTA	6039
Db	1846	TAAAGATGTTTCATTTAATCATCCATGAGAAAGTCATTTTGGAGCAAAATAGCTAGCTTTTA	1905

QY 5740 AAAAGCTTAATAAGACAGAGTGAACCTCTGTACGACCAATCTCTGCTCCTACTGCGCTT 5799  
 |||||  
 Db 85169 AAAAGCTTAATAAGACAGAGTGAACCTCTGTACGACCAATCTCTGCTCCTACTGCGCTT 85110  
 |||||  
 QY 5800 ACCCTCTCTGCGCTTACTAGTACCGCAATATTTTGGAGTCCCAAGACCTCTGTGACT 5859  
 |||||  
 Db 85109 ACCCTCTCTGCGCTTACTAGTACCGCAATATTTTGGAGTCCCAAGACCTCTGTGACT 85050  
 |||||  
 QY 5860 TACAGCTCTTAATAGCATGATTTCAATATAGCTGTAAATAAACTCTACTTATGTGTACACC 5919  
 |||||  
 Db 85049 TACAGCTCTTAATAGCATGATTTCAATATAGCTGTAAATAAACTCTACTTATGTGTACACC 84990  
 |||||  
 QY 5920 ATTTTTCCTAATTTTAAAAAATTTTCAAAAGTATGAATATATTTATTTATGTAACCTCA 5979  
 |||||  
 Db 84989 ATTTTTCCTAATTTTAAAAAATTTTCAAAAGTATGAATATATTTATTTATGTAACCTCA 84930  
 |||||  
 QY 5980 TAAAGATGTTCAATTAATCATCATGAGAAAGTCAATTTTGGAGCAAAATAGTGTCTTTA 6039  
 |||||  
 Db 84929 TAAAGATGTTCAATTAATCATCATGAGAAAGTCAATTTTGGAGCAAAATAGTGTCTTTA 84870  
 |||||  
 QY 6040 AATATTTGCATATGTGAACACATGAATGAATTCGAGCTATAAAATTTGTATGTGTT 6099  
 |||||  
 Db 84869 AATATTTGCATATGTGAACACATGAATGAATTCGAGCTATAAAATTTGTATGTGTT 84810  
 |||||  
 QY 6100 TATTTTACTTAAATAGTAAATAGTTTCTTTTCAATGAGACTGGCTGCTGATGCACCT 6159  
 |||||  
 Db 84809 TATTTTACTTAAATAGTAAATAGTTTCTTTTCAATGAGACTGGCTGCTGATGCACCT 84750  
 |||||  
 QY 6160 TGGTAATGATCATGATTAATTTCTAACTGAGATATATTTGAGATTAATGATGATTAAC 6219  
 |||||  
 Db 84749 TGGTAATGATCATGATTAATTTCTAACTGAGATATATTTGAGATTAATGATGATTAAC 84690  
 |||||  
 QY 6220 ACTCTCTCAGTACATCAAAATCATTGCGAGTATTAGAAATGACCATGAGCTAAATA 6279  
 |||||  
 Db 84689 ACTCTCTCAGTACATCAAAATCATTGCGAGTATTAGAAATGACCATGAGCTAAATA 84630  
 |||||  
 QY 6280 TGCTCAACTCTGCTTTTATTTCTTAAATGCGCAAAAAAATTTTAAAAA 6330  
 |||||  
 Db 84629 TGCTCAACTCTGCTTTTATTTCTTAAATGCGCAAAAAAATTTTAAAAA 84579  
 |||||

## RESULT 6

AC138024/c  
 LOCUS AC138024 163815 bp DNA linear PRI 01-MAR-2003  
 DEFINITION Homo sapiens chromosome 1 clone RP5-1099E6, complete sequence.  
 ACCESSION AC138024 AL33782  
 VERSION AC138024.2 GI:28626669  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
 Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and  
 Haugen, E.D.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 163815)  
 Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.  
 Direct Submission  
 Submitted (10-DEC-2002) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 3 (bases 1 to 163815)  
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
 Saenphimmachak, C., Buckley, D., Kibukawa, M., Raymond, C. and  
 Haugen, E.D.  
 Direct Submission  
 Submitted (01-MAR-2003) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 On Mar 1, 2003 this sequence version replaced gi:26291485.  
 ----- Genome Center  
 Center: University of Washington Genome Center

Center Code: UWGC  
 Web site: <http://www.genome.washington.edu>  
 Contact: [uwgctg@u.washington.edu](mailto:uwgctg@u.washington.edu)  
 Drafting Center: SC

## ----- Project Information

Center project name: chr-1  
 Center clone name: RP5-1099E6 (sc0859)

## ----- Summary Statistics

Sequencing vector: plasmid; 82% of reads  
 Sequencing vector: plasmid; 108752; 18% of reads  
 Chemistry: Dye-terminator ET; 20% of reads  
 Chemistry: Dye-terminator Big Dye; 80% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 163790 bases at least Q40  
 Consensus quality: 163815 bases at least Q30  
 Insert size: 163815; sum-of-contigs  
 Quality coverage: 8.9x in Q20 bases; sum-of-contigs

## ----- Overlapping Sequences:

5': RP4-723P6 (UWGC:sc0827) AC119429, 100473-bp overlap  
 3': RP11-239122 AL358558, 100-bp overlap

## ----- Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

## ----- Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

EcoRI			HindIII			BglII		
SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	
5829	5793	1454	1511	11401	11165			
2184	2180	449	<800	5671	5862			
8065	8081	512	<800	13046	12992			
486	<800	2814	2667	1819	1798			
106	<800	1247	1246	3902	3914			
4368	4483	4695	4872	5005	5109			
1070	1062	3373	3343	3847	3787			
3044	3127	11328	10882	13968	14281			



Qy	5440	CTTAGCTAAATCATGCGCTAACAGATAATTGGATGTAAATGTTTTTCTTTTCTTTCTCTTGCTG	5499
Db	70002	CTTAGCTAAATCATGCGCTAACAGATAATTGGATGTAAATGTTTTTCTTTTCTTTCTCTTGCTG	69943
Qy	5500	TTTTCCTTCTCTTTTCTTTTCACTGTGCAACTTAATATCTCATGTTCTATGAAGAACAATTG	5559
Db	69942	TTTTCCTTCTCTTTTCTTTTCACTGTGCAACTTAATATCTCATGTTCTATGAAGAACAATTG	69883
Qy	5560	TGGGGAAAACTAATCCACGSGAAAAGATAACTTCTCTAAGCCAGCAGCATATGCTAAAGCAA	5619
Db	69882	TGGGGAAAACTAATCCACGSGAAAAGATAACTTCTCTAAGCCAGCAGCATATGCTAAAGCAA	69823
Qy	5620	GTGAGGCTCTTGTTTCGGTCACAAAATTTAAAGGCCATAAAAAAACCCTCAGTGTTAATGTAA	5679
Db	69822	GTGAGGCTCTTGTTTCGGTCACAAAATTTAAAGGCCATAAAAAAACCCTCAGTGTTAATGTAA	69763
Qy	5680	AITTTTAATGCATATTTTTTAAAAATGCAAAATCAATGTGAAAGCACTATAAAAAATATATC	5739
Db	69762	AITTTTAATGCATATTTTTTAAAAATGCAAAATCAATGTGAAAGCACTATAAAAAATATATC	69703
Qy	5740	AAAAGCTTAAATAAAGACAGATTGAACTCTGTACACGACACAATCTGCTCCAATGGCCCTT	5799
Db	69702	AAAAGCTTAAATAAAGACAGATTGAACTCTGTACACGACACAATCTGCTCCAATGGCCCTT	69643
Qy	5800	ACCCTCTCTCTGCGCTTACTAGTACCGCAATATTTTGGAAAGTCCCACTGACCTCTGTGACT	5859
Db	69642	ACCCTCTCTCTGCGCTTACTAGTACCGCAATATTTTGGAAAGTCCCACTGACCTCTGTGACT	69583
Qy	5860	TACAGCTTCTTAATAGCATGATTTCAATATAGCTGTAAAAAAACTCTACTTATGGTACACC	5919
Db	69582	TACAGCTTCTTAATAGCATGATTTCAATATAGCTGTAAAAAAACTCTACTTATGGTACACC	69523
Qy	5920	AITTTTCCAAATTTTTTAAAAAAATTTTCAAAAGTATAAGTATATATATATGTAAACTCA	5979
Db	69522	AITTTTCCAAATTTTTTAAAAAAATTTTCAAAAGTATAAGTATATATATATGTAAACTCA	69463
Qy	5980	TAAAGATGTTTCATTTAATCATCCATGAGAAAGTCATTTTGGAGCAAAATAGCTAGTCTTTA	6039
Db	69462	TAAAGATGTTTCATTTAATCATCCATGAGAAAGTCATTTTGGAGCAAAATAGCTAGTCTTTA	69403
Qy	6040	AAATATATGCAATATGTGAAGCAATGAAATGGAAATTCGAGCTATAAAAAATTTGTAATGTTT	6099
Db	69402	AAATATATGCAATATGTGAAGCAATGAAATGGAAATTCGAGCTATAAAAAATTTGTAATGTTT	69343
Qy	6100	TATTTTTTACTTAAAAATAGTAAATAGTTTGCCTTTTCATTTGAGAGCTGGCTGCTGATGCACCT	6159
Db	69342	TATTTTTTACTTAAAAATAGTAAATAGTTTGCCTTTTCATTTGAGAGCTGGCTGCTGATGCACCT	69283
Qy	6160	TGCTTAATGAATCATGATATATATTTCAACTGCAGATATATTTGAGATTAATGCAATGATTAAT	6219
Db	69282	TGCTTAATGAATCATGATATATATTTCAACTGCAGATATATTTGAGATTAATGCAATGATTAAT	69223
Qy	6220	ACTCTCTCAGTACATCAAAATCATTTGCAGAGTATTTAGAAATTTGAACCACTTAGCTGATAAA	6279
Db	69222	ACTCTCTCAGTACATCAAAATCATTTGCAGAGTATTTAGAAATTTGAACCACTTAGCTGATAAA	69163
Qy	6280	TGCTCAACTCTCTCTTTATTTCTTAAAAATGCAAAAAAATAAAAAAAAAAAAAA	6330
Db	69162	TGCTCAACTCTCTCTTTATTTCTTAAAAATGCAAAAAAATAAAAAAAAAAAAAA	69112

RESULT 7	AC021209	184849 bp	DNA	linear	HTG 16-JUL-2000
LOCUS	AC021209				
DEFINITION	Homo sapiens chromosome 1 clone RP11-271N19, WORKING DRAFT SEQUENCE, 11 unordered pieces.				
ACCESSION	AC021209				
VERSION	AC021209.4	GI:8571797			
KEYWORDS	HTG; HTGS; PHASE1; HTGS_DRAFT.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

1 (bases 1 to 184849)  
Waterston, R.H.  
The sequence of Homo s  
Unpublished  
2 (bases 1 to 184849)

Waterston, R.H.  
Direct Submission  
Submitted (15-JAN-2000)

Submitted (15-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
On Jun 18, 2000 this sequence version replaced qi:7622510.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: H NH0271N19

----- Summary Statistics -----  
Semiannual vector: M12, 70%

Sequencing vector: pM13; 70%  
Sequencing vector: pBluescript; 30%

sequencing vector: plasmid; 30%  
Chemistry: Dye-primer ET: 70% of reads

Chemistry: 2TC-primer 51; 70% of reads  
Chemistry: Dye-terminator Big Dye; 30% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 180192 bases at least Q40

Consensus quality: 181428 bases at least Q30

Consensus quality: 182156 bases at least Q20

Insert size: 182000; agarose-fp

Insert size: 183849; sum-of-contigs  
Quality coverage: 5 14 in 920 bases: average-fn

Quality coverage: 3.14 in Q20 bases; agarose-rip

[illegible]

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 11 contigs. The true order of the sequence  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

*	1	2056:	contig	of 2056 bp in length
*	2057	2156:	gap of	unknown length
*	2157	4432:	contig	of 2276 bp in length
*	4433	4532:	gap of	unknown length
*	4533	8368:	contig	of 3836 bp in length
*	8369	8468:	gap of	unknown length
*	8469	12467:	contig	of 3999 bp in length
*	12468	12567:	gap of	unknown length
*	12568	21302:	contig	of 8735 bp in length
*	21303	21402:	gap of	unknown length
*	21403	31546:	contig	of 10144 bp in length
*	31547	31646:	gap of	unknown length
*	31647	47554:	contig	of 15908 bp in length
*	47555	47654:	gap of	unknown length
*	47655	77534:	contig	of 29880 bp in length
*	77535	77634:	gap of	unknown length
*	77635	105104:	contig	of 27470 bp in length
*	105105	105204:	gap of	unknown length
*	105205	139458:	contig	of 34254 bp in length
*	139459	139558:	gap of	unknown length
*	139559	184849:	contig	of 45291 bp in length
*	184850	184949:	gap of	unknown length

```

FEATURES
source
misc_f
misc_f
misc_f

```

1023	1016	2363	2230	9032	8951	1020	1016	15331	15830
2764	2870	4285	4490	3184	3134	524	<800	3891	3870
83	<800	1862	1935	4338	4430	1353	1320	583	<800
2856	3012	3533	3343	7186	7196	4589	4483	550	<800
2011	1970	3192	3343	3045	3020	673	<800	3872	3870
4071	3999	3729	3870	4618	4838	3419	3396	1550	1580
15	<800	1293	1246	2827	2794	1529	1479	114	<800
18676	18998	128	<800	29	<800	6982	6935	1989	2039
1216	1201	6192	6130	7295	7196	3464	3396	459	<800
4182	4323	709	<800	7254	7196	3821	3783	2991	2840
3994	3999	3355	3343	5375	5622	811	817	4471	4490
119	<800	3037	3055	5244	5280	1421	1422	1421	1422
2538	2539	4420	4490	2663	2638				
1811	1782	1526	1580	463	<800				
3123	3267	623	<800	4549	4529				
1440	1405	1206	1246	2980	2919				
637	<800	4014	4140	4004	4239				
2972	3012	31	<800	873	896				
66	<800	4933	5103	6196	6116				
2591	2620	943	843	798	823				
229	<800	244	<800	9479	9364				
1358	1320	3156	3055	1075	1071				
8081	8081	660	<800	5999	5862				
13	<800	1591	1668	325	<800				
803	817	12379	12010	5992	5862				
5206	5155	172	<800	2966	2919				
4030	3999	203	<800	3450	3373				
7174	7174	4074	4140	1220	1178				
4168	4154	452	<800	8213	8143				
892	904	1024	1082	494	<800				
9091	8928	1081	1082						
2109	2081	1072	1082						
9521	9349	4339	4490						
2653	2749	2883	2840						
11613	11225	2453	2285						
3359	3396	7821	7671						

Query Match 24.4%; Score 1546.2; DB 9; Length 163815;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1548; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	4780	AGCACGACAACTGAGGCCACCTCTGGTTAAAGGATCAACAGCACAAATCCATCT	4839
DB	70662	AGCACGACAACTGAGGCCACCTCTGGTTAAAGGATCAACAGCACAAATCCATCT	70603
QY	4840	TAAGTGGTTCCACCTGAGAACTGAATGAGCCCTCTCTCTATATATCAGCTGGAAGGAG	4899
DB	70602	TAGTGGTTCCACCTGAGAACTGAATGAGCCCTCTCTCTATATATCAGCTGGAAGGAG	70543
QY	4900	AGAGTCATCTTACCAGCTCTGATGACCCAGATGATGAAGAAATCCGTTTCATAGGAA	4959
DB	70542	AGAGTCATCTTACCAGCTCTGATGACCCAGATGATGAAGAAATCCGTTTCATAGGAA	70483
QY	4960	TGGTATTGTAATTTCCAGCTCCACTCACCAGTCAATACAGACTTCACCTGGTAAGTG	5019
DB	70482	TGGTATTGTAATTTCCAGCTCCACTCACCAGTCAATACAGACTTCACCTGGTAAGTG	70423
QY	5020	TGTTTGACATTCGTTTATTTAGGAGACAGAGCTCCAAAATGTTTCTATATTTTCATA	5079
DB	70422	TGTTTGACATTCGTTTATTTAGGAGACAGAGCTCCAAAATGTTTCTATATTTTCATA	70363
QY	5080	TCCCTTTACATGAATTTTATTTATATACCTTACCTTAGAGAAATCTAATTCAGCCCTTGAT	5139
DB	70362	TCCCTTTACATGAATTTTATTTATATACCTTACCTTAGAGAAATCTAATTCAGCCCTTGAT	70303
QY	5140	AGCTTTTCCTGATGTTTCAGCATGCTCCATCTTTTGAATTTCTGGGAAAAAAGTCAG	5199
DB	70302	AGCTTTTCCTGATGTTTCAGCATGCTCCATCTTTTGAATTTCTGGGAAAAAAGTCAG	70243
QY	5200	GTAAGTGAAGGAAAAAATAAAGATGAAGATGAAGAGCAGCCTTATTTGGATCAAA	5259
DB	70242	GTAAGTGAAGGAAAAAATAAAGATGAAGATGAAGAGCAGCCTTATTTGGATCAAA	70183
QY	5260	GTAAGTGTCTTTGTATTTCTTTTGTGAAGTATGCGAGCAATGTTCTTGAATAT	5319
DB	70182	GTAAGTGTCTTTGTATTTCTTTTGTGAAGTATGCGAGCAATGTTCTTGAATAT	70123
QY	5320	TATTCACGTCTCTCTGAGCAATGAGTTTGCAAAATGCCCTCATGCTATTTGGAGATTC	5379
DB	70122	TATTCACGTCTCTCTGAGCAATGAGTTTGCAAAATGCCCTCATGCTATTTGGAGATTC	70063
QY	5380	TCAGTATGACCCCGTTACTGAACTCCAAAAGCATTTGAAGAAAGCTATTTCAACTTTG	5439
DB	70062	TCAGTATGACCCCGTTACTGAACTCCAAAAGCATTTGAAGAAAGCTATTTCAACTTTG	70003

AUTHORS Weston, M.D., Eudy, J.D., Fugita, S., Yao, S.-F., Usami, S., Cremers, C., Greenburg, J., Ramesar, R., Martini, A., Moller, C., Smith, R.J., Sumegi, J., and Kimberling, W.J.  
TITLE Genomic structure and identification of novel mutations in usherin, the gene responsible for Usher syndrome type IIa  
JOURNAL Am. J. Hum. Genet. 66 (4), 1199-1210 (2000)  
MEDLINE 20206315  
PUBMED 10729113  
REFERENCE 2 (bases 1 to 2749)  
AUTHORS Weston, M.D., Eudy, J.D., Fugita, T., Sumegi, J., and Kimberling, W.J.  
TITLE Direct Submission  
JOURNAL Submitted (14-SEP-1998) Genetics, Boys Town National Research Hospital, 555 North 30th Street, Omaha, NE 68131, USA  
FEATURES  
source  
1. .2749  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
666. .848  
/number=1  
1522. .2210  
/number=2  
BASE COUNT 810 a 536 c 561 g 842 t  
ORIGIN  
Query Match 10.8%; Score 683.4; DB 9; Length 2749;  
Best Local Similarity 98.0%; Pred. No. 1.8e-140;  
Matches 701; Conservative 1; Mismatches 12; Indels 1; Gaps 1;  
QY 182 GATACAGCAGCTACTCATGCTCTCGCCATTGCTAAGAACGTCGTTGGTATTACCTTACT 241  
DB 1521 GATACAGCAGCTACTCATGCTCTCGCCATTGCTAAGAACGTCGTTGGTATTACCTTACT 1580  
QY 242 CTGAGAACCTGTCTGCGAGTTTCAGAAAATGAGATATCGCAATCATCACTTAAAGTACCTT 301  
DB 1581 CTGAGAACCTGTCTGCGAGTTTCAGAAAATGAGATATCGCAATCATCACTTAAAGTACCTT 1640  
QY 302 GCTTCAAGTATTGTCGCAAGTGGCGTGGCTGATATTATTATTAGAAATGCTTTATC 361  
DB 1641 GCTTCAAGTATTGTCGCAAGTGGCGTGGCTGATATTATTATTAGAAATGCTTTATC 1700  
QY 362 AGGAGGAGAAATGC - TTTTGTGAACATGAATGCCAGTCTTTTCATTGGGCTCTGGCTT 420  
DB 1701 AGGAGGAGAAATGCTTTTGTGAACATGAATGCCAGTCTTTTCATTGGGCTCTGGCTT 1760  
QY 421 CTGTGTTTCAGTCATGAATGCTGATCTTTCGCTATTTTGTCTCAATATCTTCACTGA 480  
DB 1761 CTGTGTTTCAGTCATGAATGCTGATCTTTCGCTATTTTGTCTCAATATCTTCACTGA 1820  
QY 481 GTCAAGAGTCTTTTCCCAAGCTGGAGAACGTTGGAGCTTTTCAAGAAAGTTTCCATCGT 540  
DB 1821 GTCAAGAGTCTTTTCCCAAGCTGGAGAACGTTGGAGCTTTTCAAGAAAGTTTCCATCGT 1880  
QY 541 GCCAACCCAGCAGTATGTGGATCTCCAGACCGAAGCACTTTTGTTCAGAGCTCTGCTGC 600  
DB 1881 GCCAACCCAGCAGTATGTGGATCTCCAGACCGAAGCACTTTTGTTCAGAGCTCTGCTGC 1940  
QY 601 TCGTGAAGTATTACGTTCTGTACCCAGCGGTTTGTATTCAGAGATGCCCCATACAGATC 660  
DB 1941 TCGTGAAGTATTACGTTCTGTACCCAGCGGTTTGTATTCAGAGATGCCCCATACAGATC 2000  
QY 661 TTCACACCTTACTACACGCTTCTTCTCAGCAGGCTCTCAGTCTGATCATCACACCCAGA 720  
DB 2001 TTCACACCTTACTACACGCTTCTTCTCAGCAGGCTCTCAGTCTGATCATCACACCCAGA 2060  
QY 721 CAAGATATGATTCGATCTTAAGCCCATAGCAATCTCGAAGTTTATTTTGGAAATCA 780  
DB 2061 CAAGATATGATTCGATCTTAAGCCCATAGCAATCTCGAAGTTTATTTTGGAAATCA 2120  
QY 781 CAAGAGCTGCTTTTCTTCTCTCTCTCTCCAAAGCTGATGGATCATTTACCTTAGCTGT 840  
DB 2121 CAAGAGCTGCTTTTCTTCTCTCTCTCTCCAAAGCTGATGGATCATTTACCTTAGCTGT 2180

QY 841 ATGGCTGAACCTGAGCAACGAGGTGTAATGTTGTTATAGAAAGACGCTAGAT 895  
DB 2181 ATGGCTGAACCTGAGCAACGAGGTGTAATGTTGTTATAGAAAGACGCTAGAT 2235  
RESULT 9  
AL445650/c  
LOCUS Human DNA sequence from clone RP11-152K20 on chromosome 1, complete sequence.  
DEFINITION AL445650  
ACCESSION AL445650  
VERSION AL445650.9 GI:116111403  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 153705)  
AUTHORS Martin, S.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT On Dec 9, 2000 this sequence version replaced gi:11602591.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1  
RP11-152K20 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6  
This sequence is the entire insert of clone RP11-152K20. The true right end of clone RP11-239122 is at 26243 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.  
FEATURES  
source  
1. 153705  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-152K20"  
/clone.lib="RPCI-11.1"  
3. 481  
/note="match: GSS: Em:AQ383997"  
368. .666  
/note="AluSx repeat: matches 1. .303 of consensus"  
936. .986  
/note="L2 repeat: matches 2700. .2750 of consensus"  
947. 1184  
/note="MIR repeat: matches 32. .257 of consensus"  
1181. .1740  
/note="match: GSS: Em:AQ590646"  
1185. .1838  
/note="match: GSS: Em:A2520510"  
1201. .1394  
misc\_feature  
repeat\_region  
repeat\_region  
repeat\_region  
misc\_feature  
misc\_feature  
misc\_feature



```

/note="assembly_name:Contig11
clone_end:SP6
vector_side:left"
8469..12467
/note="assembly_name:Contig12"
12568..21302
/note="assembly_name:Contig13"
21403..31546
/note="assembly_name:Contig14"
31647..47554
/note="assembly_name:Contig15"
47655..77534
/note="assembly_name:Contig16"
77635..105104
/note="assembly_name:Contig17
clone_end:T7
vector_side:left"
105205..139458
/note="assembly_name:Contig18"
139559..184849
/note="assembly_name:Contig19"
32713 C 33432 G 57811 T
BASE COUNT 59891 A
ORIGIN
```

Query Match	24.2%;	Score 1531;	DB 2;	Length 184849;
Best Local Similarity	99.6%;	Prod. No. 0;		
Matches 1545;	Conservative	0;	Mismatches 5;	Indels 1; Gaps 1;
QY	4780	AGCACGACACAACCTGAGGCCACCTCTGGTTAAAGGAATCAACAGCACACCAATCCATCT	4839	
Db	110306	AGCACGACACAACCTGAGGCCACCTCTGGTTAAAGGAATCAACAGCACACCAATCCATCT	110965	
QY	4840	TAAGTGGTTTCCACCTCGAAGAACTGAATGGACCCCTCTCTATATATCAGCTGGAAGGAG	4899	
Db	110966	TAGTGGTTTCCACCTGAAGAACTGAATGGACCCCTCTCTATATATCAGCTGGAAGGAG	111025	
QY	4900	AGAGTCATCTCTACAGCTCTGATGACCAAGATGATGAAGAAATCGGTTTCAATAGGAAA	4959	
Db	111026	AGAGTCATCTCTACAGCTCTGATGACCAAGATGATGAAGAAATCGGTTTCAATAGGAAA	111085	
QY	4960	TGGGTATTGTAAATTTTCCAGCTCCACCTCAGCTCAATACAGACTTCACCTGGTAAGTG	5019	
Db	111086	TGGGTATTGTAAATTTTCCAGCTCCACCTCAGCTCAATACAGACTTCACCTGGTAAGTG	111145	
QY	5020	TGTTTGACATTGCTTTATTTAGGAGACAGAGCTCCAAAATGTTTTCTATATTTTCATA	5079	
Db	111146	TGTTTGACATTGCTTTATTTAGGAGACAGAGCTCCAAAATGTTTTCTATATTTTCATA	111205	
QY	5080	TCCCTTTACAATGAATTTTTTATATATACCTACTTAGAGAAATACTAATTCAGCCCTTTGAT	5139	
Db	111206	TCCCTTTACAATGAATTTTTTATATATACCTACTTAGAGAAATACTAATTCAGCCCTTTGAT	111265	
QY	5140	AGCTTTTGCCTGATTGTTTCAGCATGTCCTCTTTTAGAATCTCTGGGAAAAAAGTCAG	5199	
Db	111266	AGCTTTTGCCTGATTGTTTCAGCATGTCCTCTTTTAGAATCTCTGGGAAAAAAGTCAG	111325	
QY	5200	GTAAGTGAAGGAAAGGAAAAATAAAGATGAAGATGAAGAGCAGCCCTTATTTGGATCAA	5259	
Db	111326	GTAAGTGAAGGAAAGGAAAAATAAAGATGAAGATGAAGAGCAGCCCTTATTTGGATCAA	111385	
QY	5260	GTATGTCGTTTGTAATTTGTCCTTTTGTGAAGATATGTGCCAGACATGTTCTTGAATAT	5319	
Db	111386	GTATGTCGTTTGTAATTTGTCCTTTTGTGAAGATATGTGCCAGACATGTTCTTGAATAT	111445	
QY	5320	TATTCACTGCTCTCTGAGCAATGAGTTTGCAAAATGCCCTCATGCTATTTGGAGATTC	5379	
Db	111446	TATTCACTGCTCTCTGAGCAATGAGTTTGCAAAATGCCCTCATGCTATTTGGAGATTC	111505	
QY	5380	TCAGTATGCACCCCGTTACTGAAACTCCAAAAGCATTTGTAAGAAAGCTATTTCACCTTGG	5439	
Db	111506	TCAGTATGCACCCCGTTACTGAAACTCCAAAAGCATTTGTAAGAAAGCTATTTCACCTTGG	111565	
QY	5440	CTTAGCTAAATCATGCCTAACAGATATTTGATGTAATGTTTTCTTTTTCTCTCTCTGCTG	5499	

Db	111566	CTTAGCTAATCATGCCATAACAGATATTGTAGTAATGTTTCTTTCTTTCTCTCTGCTG	111625
Qy	5500	TTTCTCTCTCTTTTTCACCTGTGACAACTTAATATCTCAATGTTCTATGAAGAACATG	5559
Db	111626	TTTCTCTCTCTTTTTCACCTGTGACAACTTAATATCTCAATGTTCTATGAAGAACATG	111685
Qy	5560	TGGGGAAAACTAAATCCAGGGAAAAAGATAAATCTCTCAAGCCAGGACTATGTAAGCAA	5619
Db	111686	TGGGGAAAACTAAATCCAGGGAAAAAGATAAATCTCTCAAGCCAGGACTATGTAAGCAA	111745
Qy	5620	GTGAGGCTCTGTTTTCGGTCAAAAAATTTAAAGGCACATAAAAACTCAGTGTATGTAA	5679
Db	111746	GTGAGGCTCTGTTTTCGGTCAAAAAATTTAAAGGCACATAAAAACTCAGTGTATGTAA	111805
Qy	5680	ATTTTAATGCAATATTTTAAAAATGAAATCAATGTGAAAGCACTATAAANAATATTATC	5739
Db	111806	ATTTTAATGCAATATTTT-AAAATGAAATCAATGTGAAAGCACTATAAANAATATTATC	111864
Qy	5740	AAAACTTAAATAAAGACAGATTGAACTCTGTACAGCACAACTCTGCCTCACTGGCCTT	5799
Db	111865	AAAACTTAAATAAAGACAGATTGAACTCTGTACAGCACAACTCTGCCTCACTGGCCTT	111924
Qy	5800	ACCTCTCTCTGGCCTTACTAGTACCGCAATATTTTGGAAAGTCCATGACCTCTGTGACT	5859
Db	111925	ACCTCTCTCTGGCCTTACTAGTACCGCAATATTTTGGAAAGTCCATGACCTCTGTGACT	111984
Qy	5860	TACAGCTTCTAAATAGCATGATTTCAAATATAGTCTGTAAAAAATCTACTTATGGTACACC	5919
Db	111985	TACAGCTTCTAAATAGCATGATTTCAAATATAGTCTGTAAAAAATCTACTTATGGTACACC	112044
Qy	5920	ATTTTTCCAATTTTAAAAAATTTACAAAGTATPAAGATATATATTATGTAAACTCA	5979
Db	112045	ATTTTTCCAATTTTAAAAAATTTACAAAGTATPAAGATATATATTATGTAAACTCA	112104
Qy	5980	TAAAGATGTTCAATTAATCATCCATGAGGAAGTCAATTTGGAGCAATAGTAGTCTTTA	6039
Db	112105	TAAAGATGTTCAATTAATCATCCATGAGGAAGTCAATTTGGAGCAATAGTAGTCTTTA	112164
Qy	6040	AAATATTGTCATATGCTGAAGACAATGAAATGGAATTCGAGCTATATAAANAATTTGATTGTT	6099
Db	112165	AAATATTGTCATATGCTGAAGACAATGGAATTCGAGCTATATAAANAATTTGATTGTT	112224
Qy	6100	TATTTTACTTTAAAAATAGTAATGTTGCTTTTTCATGTAGACTGGCTGCTGATCACT	6159
Db	112225	TATTTTACTTTAAAAATAGTAATGTTGCTTTTTCATGTAGACTGGCTGCTGATCACT	112284
Qy	6160	TGGTAATGAATCATGATTATATCTTAACCTGAGATATATGAGATTATGCAATGATTAAT	6219
Db	112285	TGGTAATGAATCATGATTATATCTTAACCTGAGATATATGAGATTATGCAATGATTAAT	112344
Qy	6220	ACTCTCTCAGTACATAAAAAATCAATTCGAGAGTATTTAGAAATTTGAACCAATGAGCTAAAAA	6279
Db	112345	ACTCTCTCAGTACATAAAAAATCAATTCGAGAGTATTTAGAAATTTGAACCAATGAGCTAAAAA	112404
Qy	6280	TGCTCAACTCTGCTTTATATATCTTAAAAATGGCAAAAAAATAAAAAA	6330
Db	112405	TGCTCAACTCTGCTTTATATATCTTAAAAATGGCAAAAAAATAAAAAA	112455

RESULT 8  
HSUSH2A01

LOCUS	HSUSH2A01	2749 bp	DNA	linear	PRI 14-NOV-2000			
DEFINITION	Homo sapiens Usher syndrome type IIA protein gene, exons 1 and 2.							
ACCESSION	AF091873							
VERSION	AF091873.1	GI:8515069						
KEYWORDS								
SEGMENT	1 of 17							
SOURCE	Homo sapiens (human)							
ORGANISM	Homo sapiens							
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
REFERENCE	1 (bases 1 to 2749)							





```
repeat_region /note="match: GSS: Em:AQ503999"
1219. .1340
repeat_region /note="61 copies 2 mer ta 64% conserved"
1229. .1340
repeat_region /note="28 copies 4 mer tata 66% conserved"
2158. .2568
repeat_region /note="MLT1C repeat: matches 54. .466 of consensus"
3375. .3617
repeat_region /note="MER58A repeat: matches 1. .223 of consensus"
4003. .4172
misc_feature /note="LiMC1 repeat: matches 6138. .6306 of consensus"
complement(4777. .5298)
misc_feature /note="match: GSS: Em:AQ525021"
5923. .6252
misc_feature /note="match: GSS: Em:AQ152396"
complement(5934. .6086)
repeat_region /note="match: GSS: Em:AQ186738"
7304. .7337
repeat_region /note="MIR repeat: matches 229. .262 of consensus"
7338. .7648
repeat_region /note="AluSg repeat: matches 1. .301 of consensus"
7649. .7765
repeat_region /note="MIR repeat: matches 111. .230 of consensus"
7784. .8093
repeat_region /note="AluSx repeat: matches 1. .309 of consensus"
8095. .8409
misc_feature /note="AluS1 repeat: matches 1. .304 of consensus"
complement(10099. .10432)
repeat_region /note="match: GSS: Em:AQ529802 Em:AQ530310"
11036. .11169
repeat_region /note="L2 repeat: matches 1170. .1301 of consensus"
11235. .11506
repeat_region /note="LiMC4 repeat: matches 7574. .7834 of consensus"
11507. .11964
repeat_region /note="LiPA4 repeat: matches 5689. .6146 of consensus"
11965. .12330
repeat_region /note="LiMC4 repeat: matches 7172. .7574 of consensus"
12351. .12698
repeat_region /note="LiMC4 repeat: matches 6643. .6996 of consensus"
12870. .13925
repeat_region /note="TiGSR1 repeat: matches 55. .1118 of consensus"
15429. .15642
repeat_region /note="107 copies 2 mer aa 72% conserved"
15442. .15637
misc_feature /note="49 copies 4 mer gaaa 82% conserved"
16549. .17186
misc_feature /note="match: GSS: Em:AQ239802"
17713. .18197
repeat_region /note="match: GSS: Em:AQ613916"
19478. .19712
misc_feature /note="MER8 repeat: matches 11. .239 of consensus"
complement(19904. .20377)
repeat_region /note="match: GSS: Em:AQ885526"
21383. .21707
repeat_region /note="AluJb repeat: matches 1. .302 of consensus"
21843. .21905
misc_feature /note="LiPA8 repeat: matches 6101. .6163 of consensus"
23594. .24161
misc_feature /note="match: GSS: Em:AQ377092"
23595. .24035
repeat_region /note="match: GSS: Em:AQ507156"
23923. .24218
misc_feature /note="AluY repeat: matches 1. .308 of consensus"
24217. .24225
/note="Sequence from overlapping clone ba239122
(AL358858). Assembly confirmed by restriction digest
data."
repeat_region 24249. .24272
/note="12 copies 2 mer ta 100% conserved"
24274. .24574
repeat_region /note="AluY repeat: matches 2. .294 of consensus"
complement(25249. .25594)
misc_feature /note="match: GSS: Em:AQ551687"
26018. .26048

complement(25692. .26243)
/note="match: GSS: Em:AQ561898"
29416. .29473
/note="MIR repeat: matches 88. .146 of consensus"
29658. .29689
/note="MIR repeat: matches 106. .137 of consensus"
29924. .30224
/note="AluY repeat: matches 1. .307 of consensus"
31517. .32047
/note="LiMC/D repeat: matches 5154. .5658 of consensus"
32428. .32724
/note="AluY repeat: matches 1. .297 of consensus"
33358. .33581
/note="MER20 repeat: matches 8. .213 of consensus"
35048. .35677
/note="MLT2D repeat: matches 1. .548 of consensus"
36085. .36235
/note="LiMD repeat: matches 232. .380 of consensus"
36236. .36275
/note="20 copies 2 mer ca 87% conserved"
36239. .36278
/note="10 copies 4 mer acac 87% conserved"
37305. .37373
/note="LiMD repeat: matches 542. .623 of consensus"
37651. .38715
/note="LiMD repeat: matches 980. .2053 of consensus"
39119. .39615
/note="LiM4 repeat: matches 2110. .2596 of consensus"
39808. .40105
/note="LiM4 repeat: matches 2679. .2947 of consensus"
40106. .40535
/note="TiGSR2a repeat: matches 2. .431 of consensus"
40536. .40614
/note="LiM4 repeat: matches 2947. .3030 of consensus"
40660. .40740
/note="LiM4 repeat: matches 3016. .3101 of consensus"
40731. .41438
/note="LiMD2 repeat: matches 3750. .4490 of consensus"
41439. .41818
/note="TH21C repeat: matches 1. .371 of consensus"
41819. .42252
/note="LiMD2 repeat: matches 4490. .4914 of consensus"
42326. .42630
/note="AluJb repeat: matches 1. .303 of consensus"
42662. .42862
/note="AluJo repeat: matches 1. .220 of consensus"
42877. .43009
/note="LiMD2 repeat: matches 5017. .5153 of consensus"
43039. .44071
/note="LiMD2 repeat: matches 5290. .6335 of consensus"
44769. .44844
/note="19 copies 4 mer ctic 75% conserved"
46060. .46764
/note="MER44C repeat: matches 1. .728 of consensus"
46827. .48181
/note="LiP4 repeat: matches 4387. .5765 of consensus"
48240. .48346
/note="MER5B repeat: matches 60. .177 of consensus"
48757. .48912
/note="L2 repeat: matches 2168. .2335 of consensus"
49037. .49302

Query Match 10.8%; Score 683.4; DB 9; Length 153705;
Best Local Similarity 98.0%; Pred No. 1.8e-140;
Matches 701; Conservative 1; Mismatches 13; Indels 1; Gaps 1;

QY 182 GATACACGACTCTCATGTCTTGCCATTGCTGAAGACGCTGTGGTATTACCTTACT 241
|||||
DB 60248 GATACACGACTCTCATGTCTTGCCATTGCTGAAGACGCTGTGGTATTACCTTACT 60189
|||||
QY 242 CTGAGAACGTCTGTCAGTTTCCAGAAATGAGTATCGCAACATCACTTAAAGTACCCT 301
|||||
DB 60188 CTGAGAACGTCTGTCAGTTTCCAGAAATGAGTATCGCAACATCACTTAAAGTACCCT 60129
|||||
```

KEYWORDS  
SOURCE  
ORGANISM  
  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 290708)  
McLay, K.  
Direct Submission  
Submitted (16-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Sep 27, 2002 this sequence version replaced gi:22415846.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: dj865G23  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: M13; M7815; 0% of reads  
Chemistry: Dye-terminator ABI; 99% of reads  
Chemistry: Dye-terminator ABI; 0% of reads  
Chemistry: Dye-terminator; 21% of reads  
Chemistry: Dye-terminator ET-amersham; 9% of reads  
Dye-terminator Big Dye; 68% of reads  
Consensus quality: 276213 bases at least Q40  
Consensus quality: 280417 bases at least Q30  
Consensus quality: 282780 bases at least Q20  
Insert size: 285208; sum-of-contigs  
Insert size: 134879; 2.4% error; agarose-fp  
Quality coverage: 6.08x in Q20 bases; sum-of-contigs Quality  
coverage: 15.17x in Q20 bases; agarose-fp

## COMMENT

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 56 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1. 2509: contig of 2509 bp in length  
2510: gap of 100 bp  
2610: 9943: contig of 7334 bp in length  
9944: 10043: gap of 100 bp  
10044: 12149: contig of 2106 bp in length  
12150: 12249: gap of 100 bp  
12250: 15904: contig of 3655 bp in length  
15905: 16004: gap of 100 bp  
16005: 20365: contig of 4561 bp in length  
20366: 20665: gap of 100 bp  
20666: 25474: contig of 4809 bp in length  
25475: 25574: gap of 100 bp  
25575: 27837: contig of 2263 bp in length  
27838: 27937: gap of 100 bp  
27939: 30564: contig of 2627 bp in length  
30565: 30664: gap of 100 bp  
30666: 34794: contig of 4130 bp in length  
34795: 34894: gap of 100 bp  
34895: 52088: contig of 17194 bp in length  
52089: 52188: gap of 100 bp  
52189: 58992: contig of 6804 bp in length  
58993: 59092: gap of 100 bp  
59093: 63330: contig of 4238 bp in length  
63331: 63430: gap of 100 bp  
63431: 65806: contig of 2376 bp in length  
65807: 65906: gap of 100 bp  
65907: 69790: contig of 3884 bp in length  
69791: 69890: gap of 100 bp  
69891: 77440: contig of 7550 bp in length  
77441: 77540: gap of 100 bp

77541: 84031: contig of 6491 bp in length  
84032: 84131: gap of 100 bp  
84132: 91442: contig of 7311 bp in length  
91443: 91542: gap of 100 bp  
91543: 94381: contig of 2839 bp in length  
94382: 94481: gap of 100 bp  
94482: 94811: contig of 3700 bp in length  
94812: 98281: gap of 100 bp  
98282: 102419: contig of 4138 bp in length  
102420: 102519: gap of 100 bp  
102520: 102520: contig of 4779 bp in length  
102521: 107398: gap of 100 bp  
107399: 110315: contig of 2917 bp in length  
110316: 110415: gap of 100 bp  
110416: 115743: contig of 5328 bp in length  
115744: 115843: gap of 100 bp  
115844: 119111: contig of 3268 bp in length  
119112: 119211: gap of 100 bp  
119212: 121219: contig of 2008 bp in length  
121220: 121319: gap of 100 bp  
121320: 123497: contig of 2178 bp in length  
123498: 123597: gap of 100 bp  
123598: 130224: contig of 6627 bp in length  
130225: 130324: gap of 100 bp  
130325: 135208: contig of 4884 bp in length  
135209: 135308: gap of 100 bp  
135309: 142309: contig of 7001 bp in length  
142310: 142409: gap of 100 bp  
142410: 144638: contig of 2229 bp in length  
144639: 144738: gap of 100 bp  
144739: 155510: contig of 10772 bp in length  
155511: 155610: gap of 100 bp  
155611: 160588: contig of 4978 bp in length  
160589: 160688: gap of 100 bp  
160689: 167138: contig of 6450 bp in length  
167139: 167238: gap of 100 bp  
167239: 172564: contig of 5326 bp in length  
172565: 172664: gap of 100 bp  
172665: 176139: contig of 3475 bp in length  
176140: 176239: gap of 100 bp  
176240: 178634: contig of 2395 bp in length  
178635: 178734: gap of 100 bp  
178735: 181546: contig of 2812 bp in length  
181547: 181646: gap of 100 bp  
181647: 183844: contig of 2198 bp in length  
183845: 183945: gap of 100 bp  
183946: 191043: contig of 7099 bp in length  
191044: 191143: gap of 100 bp  
191144: 195461: contig of 4318 bp in length  
195462: 195561: gap of 100 bp  
195562: 201981: contig of 6420 bp in length  
201982: 202081: gap of 100 bp  
202082: 204238: contig of 2157 bp in length  
204239: 204338: gap of 100 bp  
204339: 215768: contig of 11430 bp in length  
215769: 215868: gap of 100 bp  
215869: 218007: contig of 2139 bp in length  
218008: 218107: gap of 100 bp  
218109: 223568: contig of 5461 bp in length  
223569: 223668: gap of 100 bp  
223669: 226437: contig of 2769 bp in length  
226438: 226537: gap of 100 bp  
226538: 229262: contig of 2725 bp in length  
229263: 229362: gap of 100 bp  
229363: 229363: contig of 5869 bp in length  
229364: 235231: gap of 100 bp  
235232: 235331: contig of 2251 bp in length  
235333: 237582: gap of 100 bp  
237583: 237682: contig of 24848 bp in length  
237683: 262530: gap of 100 bp  
262531: 262630: contig of 7173 bp in length  
262631: 269803: gap of 100 bp  
269804: 272850: contig of 2947 bp in length

## JOURNAL

Submitted (15-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 On Jun 18, 2000 this sequence version replaced gi:7622510.

## COMMENT

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc/index.shtml>  
 ----- Project Information -----  
 Center project name: H NH0271N19  
 ----- Summary Statistics -----  
 Sequencing vector: M13; 70%  
 Chemistry: Dye-terminator ET; 30% of reads  
 Chemistry: Dye-terminator Big Dye; 30% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 180192 bases at least Q40  
 Consensus quality: 181428 bases at least Q30  
 Consensus quality: 182156 bases at least Q20  
 Insert size: 182000; agarose-fp  
 Insert size: 183849; sum-of-contigs  
 Quality coverage: 5.14 in Q20 bases; agarose-fp  
 Quality coverage: 4.97 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2056: contig of 2056 bp in length  
 2057 2156: gap of unknown length  
 2157 4432: contig of 2276 bp in length  
 4433 4532: gap of unknown length  
 4533 8368: contig of 3836 bp in length  
 8369 8468: gap of unknown length  
 8469 12467: contig of 3999 bp in length  
 12468 12567: gap of unknown length  
 12568 21302: contig of 8735 bp in length  
 21303 31546: gap of unknown length  
 31547 31647: gap of unknown length  
 31647 47554: contig of 15908 bp in length  
 47555 47654: gap of unknown length  
 47655 77535: contig of 29880 bp in length  
 77535 77635: gap of unknown length  
 77635 105104: contig of 27470 bp in length  
 105105 105204: gap of unknown length  
 105205 139458: contig of 34254 bp in length  
 139459 139558: gap of unknown length  
 139559 184849: contig of 45291 bp in length.

## FEATURES

## source

1. 184849  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-271N19"  
 1. 2056  
 /note="assembly\_name:Contig8"  
 2157. 4432  
 /note="assembly\_name:Contig10"  
 4533. 8368  
 /note="assembly\_name:Contig11"  
 clone\_end:SP6  
 vector\_side:left  
 8469. 12467  
 /note="assembly\_name:Contig12"  
 12568. 21302  
 /note="assembly\_name:Contig13"

## misc\_feature

21403..31546  
 /note="assembly\_name:Contig14"

## misc\_feature

31647..47554  
 /note="assembly\_name:Contig15"

## misc\_feature

47655..77534  
 /note="assembly\_name:Contig16"

## misc\_feature

77635..105104  
 /note="assembly\_name:Contig17"

## misc\_feature

clone\_end:T7  
 vector\_side:left  
 105205..139458

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig18"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## misc\_feature

139559..184849  
 /note="assembly\_name:Contig19"

## RESULT 12

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

AL161641 290708 bp DNA linear HTG-23-OCT-2002  
 Homo sapiens chromosome 1 clone RP5-865G23 map p35.3-36.12, \*\*\*  
 SEQUENCING IN PROGRESS \*\*\*, 56 unordered pieces.  
 AL161641  
 AL161641.20 GI:23337111

Db 526 AGACCAACCAATGTGCTGTTCAACAGGAGTGTAGTGTCACTTACAGACCAAAACAG 585  
Qy 3822 GGGTCCGAGAGGAACTTGACTTTAAGTTATATCAATTCCTATTCGCTCAGACTCTGGA 3881  
Db 586 GGGTCCGAGAGGAACTTGACTTTAAGTTATATCAATTCCTATTCGCTCAGACTCTGGA 645  
Qy 3882 CACTTACCTGGACAACTCTCAATCAATCTGCTGCTCCATAGAGAAATATATTTGCT 3941  
Db 646 CACTTACCTGGACAACTCTCAATCAATCTGCTGCTCCATAGAGAAATATATTTGCT 705  
Qy 3942 GTGCCCTTTGGCTGGTCTGAGCCATGTGTTCTTCCAGAAAGTCAATGAACCTCAGCTA 4001  
Db 706 GTGCCCTTTGGCTGGTCTGAGCCATGTGTTCTTCCAGAAAGTCAATGAACCTCAGCTA 765  
Qy 4002 CCATCTGAACTGCTGCTTCCATTTGCAAGTACATGTTTCTGTAAGGGGTGACTAGCG 4061  
Db 766 CCATCTGAACTGCTGCTTCCATTTGCAAGTACATGTTTCTGTAAGGGGTGACTAGCG 825  
Qy 4062 GGGGCTGTTTACACAGCTTGGCCATTTACAGTCAACACAGCCAGCCCTCCCAAGAC 4121  
Db 826 GGGGCTGTTTACACAGCTTGGCCATTTACAGTCAACACAGCCAGCCCTCCCAAGAC 885  
Qy 4122 TAAGTCCACCTAAGATGAGAAATCAGTTTCTACAGAACTTCAATGAGTGGTCTCCAC 4181  
Db 886 TAAGTCCACCTAAGATGAGAAATCAGTTTCTACAGAACTTCAATGAGTGGTCTCCAC 945  
Qy 4182 CAGCGGAACCTAAGTGAATATTAAG 4209  
Db 946 CAGCGGAACCTAAGTGAATATTAAG 973

RESULT 14  
AX347325/c  
LOCUS AX347325 2749 bp DNA linear PAT 01-FEB-2002  
DEFINITION Sequence 2396 from Patent WO0200928.  
ACCESSION AX347325  
VERSION AX347325.1 GI:18495213

KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1  
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.  
TITLE Diagnosis of diseases associated with the immune system  
JOURNAL Patent: WO 0200928-A 2396 03-JAN-2002;  
Epigenomics AG (DE)

FEATURES  
source  
Location/Qualifiers  
1. 2749  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 842 a 25 c 536 g 1346 t  
ORIGIN

Query Match 7.5%; Score 476.4; DB 6; Length 2749;  
Best Local Similarity 80.6%; Pred. No. 1.3e-94;  
Matches 569; Conservative 0; Mismatches 136; Indels 1; Gaps 1;

Qy 183 ATACGACGAGTACTCATGCTCTTCCCAATGCTAAGAACGCTGTTGGTATTACCTTACTC 242  
Db 1228 ATACGACGAGTACTCATGCTCTTCCCAATGCTAAGAACGCTGTTGGTATTACCTTACTC 1169  
Qy 243 TGAGAACGTGCTGAGTTTCCAGAAATGGAGTATCGCAACATCATTAAAGTACCCTG 302  
Db 1168 TAAACGCTATCTACAAATTTCCAAATAAATAATCGCAACATCATTAAAGTACCCTA 1109  
Qy 303 CTTCAGAAATATGCTGGCAAGTGGGCTGATTTATTTAGAAATGCTTTATCA 362  
Db 1108 CTTCAGAAATATGCTGGCAAGTGGGCTGATTTATTTAGAAATGCTTTATCA 1049  
Qy 363 GGAGGAGAAATGC-TTTTGTGTAACATGAATTCGCCAGTCTTTCATTTGGGCTCTGGCTC 421

Db 1048 AAAAAAATACTTTTTTATAAACAATAAATACCAATTCCTTCATTAACCTCTAACTTC 989  
Qy 422 TTGTTTCAGGTCATTGAATCTTTCCTTATTTGCTTCAATATCTTCAATATCTTCACTGAG 481  
Db 988 TTAATTCATATCAATTAATATTAATCTTTACCTATTTACTTCATATCTTCACTTAA 929  
Qy 482 TCAGAGGTCTTTTCCAGGCTGGAGAACGTTGGAGCTTTCAAGAAATGTTCCATCGTG 541  
Db 928 TCAGAAATCTTTTCCCAAAACGTAATAAATCTTTCAAAAAATTTCCATCGTA 869  
Qy 542 CCAACCCAGCAGTATGTGGACTCCAGACCGAGAGCACTTTTGTGCACAGCTCTGCTCT 601  
Db 868 CCAACCCAGCAGTATGTGGACTCCAGACCGAGAGCACTTTTGTGCACAGCTCTGCTCT 809  
Qy 602 GCTCAAGATATTCAGTTCTGTACCCAGCGGTTTGTATTCAGGATTTGCCCATACAGATCT 661  
Db 808 ACTAAAAATATTCATTTCTATACCAACGATTTTATATTCAAATTTACCCATACAACT 749  
Qy 662 TCACACCTTACCTACATGCGCTTTTCTCAGCAGGCTCAGTAGCTGCATCACACGAC 721  
Db 748 TCACACCTTACCTACATGCGCTTTTCTCAGCAGGCTCAGTAGCTGCATCACACGAC 689  
Qy 722 AAGATGATCTGCATCTTAAGCGCCATAGCAATTTCTCAAGTTTTATTTTGGAAATCAC 781  
Db 688 AAAAAATATCTACATCTTACCGCCATACAAATTTCTACAAATTTATTTTAAAAATCAC 629  
Qy 782 AAGAGCTGCTTTTCTCTCTCTTCTCCTTCTCCTTCTCCTTCTCCTTCTCCTTCTCCT 841  
Db 628 AAAAACTACTTTTCTCTCTCTTCTCCTTCTCCTTCTCCTTCTCCTTCTCCTTCTCCT 569  
Qy 842 TGGCTGAACCTGAGCAACAGGTTAATGTGTTATATAGAAAAAGA 887  
Db 568 TAACTAAACCTAAACCAACAAATAATAATAATAATAATAATAATAATAATAATAATA 523

RESULT 15  
AX347324  
LOCUS AX347324  
DEFINITION Sequence 2395 from Patent WO0200928.  
ACCESSION AX347324  
VERSION AX347324.1 GI:18495212

KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1

AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.  
TITLE Diagnosis of diseases associated with the immune system  
JOURNAL Patent: WO 0200928-A 2395 03-JAN-2002;  
Epigenomics AG (DE)

FEATURES  
Location/Qualifiers  
1. 2749  
source

BASE COUNT 810 a 25 c 561 g 1353 t  
ORIGIN

Query Match 6.9%; Score 437; DB 6; Length 2749;  
Best Local Similarity 76.5%; Pred. No. 6.9e-86;  
Matches 547; Conservative 1; Mismatches 166; Indels 1; Gaps 1;

Qy 182 GATACGACGAGTACTCATGCTCTTCCCAATGCTAAGAACGCTGTTGGTATTACCTTACT 241  
Db 1521 GATATGAGTATTTATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1580  
Qy 242 CTGAGAACGTGCTGAGTTTCCAGAAATGGAGTATCGCAACATCATTAAAGTACCCT 301  
Db 1581 TTGAGAACGTGCTGAGTTTCCAGAAATGGAGTATCGCAACATCATTAAAGTACCCT 1640  
Qy 302 GCTTCAAGATTTGCTGGCAAGTGGGCTGATTTATTTATTTAGAAATGCTTTATC 361

\* 272851 272950: gap of 100 bp  
 \* 272951 276185: contig of 3235 bp in length  
 \* 276186 276285: gap of 100 bp  
 \* 276286 278712: contig of 2427 bp in length  
 \* 278713 278812: gap of 100 bp  
 \* 278813 287208: contig of 8396 bp in length  
 \* 287209 287308: gap of 100 bp  
 \* 287309 290708: contig of 3400 bp in length.

## FEATURES

source  
 1. .290708  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosomes="1"  
 /map="p35.3-36.12"  
 /clone="RP5-865G23"  
 /clone\_lib="RPCI-5"  
 misc\_feature  
 1. .2509  
 /note="assembly\_fragment:00941  
 fragment chain:1"  
 2610. .9943  
 /note="assembly\_fragment:01476  
 fragment chain:1"  
 10044. .12149  
 /note="assembly\_fragment:01741  
 fragment chain:2"  
 12250. .15904  
 /note="assembly\_fragment:05459  
 fragment chain:2"  
 16005. .20565  
 /note="assembly\_fragment:00128"  
 20666. .25474  
 /note="assembly\_fragment:00192"  
 25575. .27837  
 /note="assembly\_fragment:00271"  
 27938. .30564  
 /note="assembly\_fragment:00358"  
 30665. .34794  
 /note="assembly\_fragment:00495"  
 34895. .52088  
 /note="assembly\_fragment:00528"  
 52189. .58992  
 /note="assembly\_fragment:00572"  
 59093. .63330  
 /note="assembly\_fragment:00579"  
 63431. .65806  
 /note="assembly\_fragment:00791"  
 65907. .69790  
 /note="assembly\_fragment:01314.0"

Query Match 10.2%; Score 645.8; DB 2; Length 290708;  
 Best Local Similarity 99.7%; Pred. No. 3.8e-132;  
 Matches 647; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2548 TATTGGGCTTAGGTGATGATGCAATTTTGGATTTAAATTTCTCCGAAGCTTTAAATGA 2607  
 DB 276369 TTTAGGGCTTAGGTGATGATGCAATTTTGGATTTAAATTTCTCCGAAGCTTTAAATGA 276428  
 QY 2608 TGTGGATGTAGCCCTGCCAGTGAACCTCCATGGCTCAGTGAACAAATTCGCAATCC 2667  
 DB 276429 TGTGGATGTAGCCCTGCCAGTGAACCTCCATGGCTCAGTGAACAAATTCGCAATCC 276488  
 QY 2668 TCACCTCGGGAGTGTAGTGCMAAAAGAACCCAAAGCACTTCACTGTGACCACTGCAG 2727  
 DB 276489 TCACCTCGGGAGTGTAGTGCMAAAAGAACCCAAAGCACTTCACTGTGACCACTGCAG 276548  
 QY 2728 AGAAACCTTTTATGGTTAGATGTCACCAATTTGTAAGCCCTGTGACTGTGACACAGCTGG 2787  
 DB 276549 AGAAACCTTTTATGGTTAGATGTCACCAATTTGTAAGCCCTGTGACTGTGACACAGCTGG 276608  
 QY 2788 ATCCCTCCCTGGGACTGTCTGTAATGCTAAGACAGGGCAGTGCATCTCGAAGCCCAATGT 2847  
 DB 276609 ATCCCTCCCTGGGACTGTCTGTAATGCTAAGACAGGGCAGTGCATCTCGAAGCCCAATGT 276668

QY 2848 TGAAGGGACAGTGCATTAATAATGTTTGAGGGAAACTTCTACCTACGCGAAAATAATTC 2907  
 DB 276669 TGAAGGGACAGTGCATTAATAATGTTTGAGGGAAACTTCTACCTACGCGAAAATAATTC 276728  
 QY 2908 TTTCTCTGTCTGCTTGCACACTGTGATAAGACTGGGACAAATAAATGGCTCTCTGCTG 2967  
 DB 276729 TTTCTCTGTCTGCTTGCACACTGTGATAAGACTGGGACAAATAAATGGCTCTCTGCTG 276788  
 QY 2968 TAAACAAATCAACAGGACAAATGCTTGCACAAATAGGGGTAAACAGGTCTTCGCTGTAATCA 3027  
 DB 276789 TAAACAAATCAACAGGACAAATGCTTGCACAAATAGGGGTAAACAGGTCTTCGCTGTAATCA 276848  
 QY 3028 GTGTGAGCTCACAGGTACCAATTTGACCAATTTGACCAATTTTCAACACTGCAGATGTGCA 3087  
 DB 276849 GTGTGAGCTCACAGGTACCAATTTGACCAATTTTCAACACTGCAGATGTGCA 276908  
 QY 3088 GTGTGATTCCTTGGGGACATTTACCTGGGACCAATTTGTGACCCCAATCAGTGGCCAGTGCCT 3147  
 DB 276909 GTGTGATTCCTTGGGGACATTTACCTGGGACCAATTTGTGACCCCAATCAGTGGCCAGTGCCT 276968  
 QY 3148 GTGTGCTTAATCTCAAGGAAGAGGTGTAAATCAGTGTCAACAGGT 3196  
 DB 276969 GTGTGCTTAATCTCAAGGAAGAGGTGTAAATCAGTGTCAACAGGT 277017

RESULT 13  
HSUSH2A14

LOCUS 2840 bp DNA linear PRI 14-NOV-2000  
 DEFINITION Homo sapiens Usher syndrome type IIA protein gene, exons 17 and 18.  
 ACCESSION AF091886  
 VERSION AF091886.1 GI:8515082  
 KEYWORDS  
 SOURCE 14 of 17  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 2840)  
 AUTHORS Weston,M.D., Eudy,J.D., Fugita,S., Yao,S.-P., Usami,S., Cremers,C.,  
 Greenburg,J., Ramesar,R., Martini,A., Moller,C., Smith,R.J.,  
 Sumegi,J. and Kimberling,W.J.  
 TITLE Genomic structure and identification of novel mutations in usherin,  
 the gene responsible for Usher syndrome type IIA  
 JOURNAL Am. J. Hum. Genet. 66 (4), 1199-1210 (2000)  
 MEDLINE 20206315  
 PUBMED 10729113  
 REFERENCE 2 (bases 1 to 2840)  
 AUTHORS Weston,M.D., Eudy,J.D., Fugita,T., Sumegi,J. and Kimberling,W.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-SEP-1998) Genetics, Boys Town National Research  
 Hospital, 555 North 30th Street, Omaha, NE 68131, USA

FEATURES  
source

1. .2840  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 466. .960  
 /number=17  
 2004. .2273  
 /number=18

BASE COUNT 944 a 502 c 476 g 917 t 1 others  
 ORIGIN

Query Match 7.9%; Score 498.4; DB 9; Length 2840;  
 Best Local Similarity 98.8%; Pred. No. 1.7e-99;  
 Matches 502; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 3702 GTATTCAATCTTCTTAGACACAGACCTGTGTACCATATACCAAAATATTCCTATTACATTG 3761  
 DB 466 GTATTCAATCTTCTTAGACACAGACCTGTGTACCATATACCAAAATATTCCTATTACATTG 525  
 QY 3762 AGACCACCAATGTGATGGTTCAACAGGAGTGTAGTGTCTCACTTACAAAGCAAAACACG 3821

[illegible]

Search completed: January 1, 2004, 11:23:45  
Job time : 14855 secs